Biblioteconomia e scienza dell'informazione

FRANCESCA TOMASI

ORGANIZZARE LA CONOSCENZA: DIGITAL HUMANITIES E WEB SEMANTICO

EDITRICE BIBLIOGRAFICA

Ci sono alcuni temi che delineano le Digital Humanities (DH) nel contesto di un Web che attribuisce alla semantica esplicita il ruolo di strumento per la gestione della conoscenza acquisibile dai dati culturali.Dopo un'analisi della storia della Humanities Computing e del suo incontro con il Web Semantico, il volume racconta l'evoluzione di questo connubio attraverso alcuni temi significativi. Partendo dallo stretto rapporto fra le DH e la Library and Information Science, e insistendo sulla componente ermeneutica delle DH, il ragionamento si alimenta del concetto di contesto, di tradizione archivistica, come sistema per l'arricchimento dei dati culturali. E dalla nozione di entità complesse e stratificate il volume arriva al tema dell'organizzazione della conoscenza, tradotto in una speciale realizzazione del Web semantico, ovvero i Linked Open Data (LOD), la tipica espressione dell'orientamento concettuale delle scienze del libro e del documento. La naturale propensione alla progettualità nelle DH, il workflow dei processi che qualificano la ricerca di settore e i nuovi mestieri del Web semantico per il digital humanist, chiudono le riflessioni.

Francesca Tomasi è professoressa associata di Archivistica, Biblioteconomia e Bibliografia presso l'Università di Bologna, dove si occupa di Digital Humanities. È stata presidente dell'Associazione italiana di Informatica Umanistica e Cultura Digitale e della Biblioteca di Discipline Umanistiche dell'Università di Bologna. È ad oggi coordinatrice del corso di laurea in Digital Humanities and Digital Knowledge e direttrice dell'Advanced Research Center for Digital Humanities dell'Università di Bologna.

BIBLIOTECONOMIA
E SCIENZA DELL'INFORMAZIONE

39.

Francesca Tomasi

ORGANIZZARE LA CONOSCENZA: DIGITAL HUMANITIES E WEB SEMANTICO

Un percorso tra archivi, biblioteche e musei

EDITRICE BIBLIOGRAFICA

Copertina: MoskitoDesign - Varese

Realizzazione eBook: CreaLibro di Davide Moroni

ISBN: 978-88-9357-478-5

Copyright © 2022 Editrice Bibliografica

Via Lesmi 6 - 20123 Milano

Proprietà letteraria riservata

https://doi.org/10.53134/9788893573573

INDICE

Introduzione

1. La humanities computing incontra il web (semantico)

1.1 Una premessa

1.2 Cronistoria

1.3 La rivoluzione del Web

1.4 La semantica (del Web) e le humanities

2. Descrizione o interpretazione? Il ruolo delle DH nella sfida ermeneutica

2.1 Digital Humanities e Library and Information Science

2.2 Descrivere e standardizzare

2.3 Interpretare

2.4 Le nuove prospettive dell'integrazione dei saperi

3. Contesti. Ovvero la lezione degli archivi

3.1 Perché i contesti?

3.2 Soggetto produttore. La funzione del ***creator***

3.3 *Provenance*. Documentare i processi

3.4 Dall'albero al grafo. Il ruolo di RiC

4. Ragionare a livelli oltre l'universo bibliografico

4.1 Livelli ed entità

4.2 Un riuso trasversale: “FRBR all’Opera”

4.3 Un modello concettuale si trasforma: verso LRMoo

5. Organizzare la conoscenza con i LOD

5.1 Cosa significa organizzazione della conoscenza?

5.2 Dati aperti e interconnessi

5.3 Visualizzare i dati. Le nuove frontiere orientate all’utente

6. Progettare risorse digitali nel cultural heritage

6.1 DH e lavoro a progetto

6.2 Buone pratiche: un *workflow* di progetto

6.3 Il ciclo vitale degli oggetti culturali

Bibliografia

INTRODUZIONE

Parlare di Digital Humanities (da ora DH), e pensare contestualmente al Web semantico, è parere ampiamente condiviso dagli studiosi che hanno dedicato al tema del rapporto fra scienze umane e discipline informatiche la propria attività di ricerca. Il Web semantico ha saputo fornire una nuova prospettiva di approccio al sapere – a livello teorico, metodologico e tecnico – che le DH hanno immediatamente sposato, perché capace di tradurre ciò che ha qualificato, fino dalle origini, il lavoro degli esperti di DH: riflettere sui sistemi computazionali per acquisire (nuova) conoscenza da dati e documenti. Esplorare prima di tutto una testualità, esito della sedimentazione degli oggetti analogici della tradizione culturale, per lanciare una sfida ermeneutica alla macchina. E se l'obiettivo, che si intende qui perseguire, è analizzare criticamente questa tradizione culturale, lo stretto legame delle DH con il tema della valorizzazione degli oggetti del patrimonio archivistico, bibliografico e museale è senz'altro naturale.

Come vedremo oltre, definire il concetto di DH non è impresa facile. Lo statuto di questo ambito, o anche settore o ancora area di ricerca, varia al variare del punto di vista dell'osservatore. Ma non ci si vorrà qui esimere da una definizione, che emerge dalla storia della disciplina e che trova la sua naturale sistematizzazione nell'esperienza condotta in Italia, attraverso l'Associazione AIUCD,[1] a livello europeo con la comunità di EADH,[2] o ancora a livello internazionale con ADHO.[3]

Ecco allora che ci sono alcuni temi ricorrenti, trasversali tanto rispetto alle discipline umanistiche coinvolte, quanto alle metodologie computazionali adottate, che contribuiscono a delineare le DH nel contesto di un Web che è quello che a oggi viviamo e che meglio qualifica l'orizzonte di ricerca delle DH, ovvero il Web semantico.

Dopo un'analisi della storia del rapporto fra Humanities Computing e Semantic Web, che rappresenta il primo contributo del

presente volume, una serie di riflessioni vogliono raccontare l'evoluzione di questo connubio attraverso alcuni temi significativi.

Primo fra tutti, lo stretto rapporto fra DH e scienze dell'informazione, ovvero quelle discipline che si occupano di elaborare strategie per l'analisi, la raccolta, la classificazione, la manipolazione, la conservazione, il reperimento e la disseminazione dell'informazione. Ecco allora che si delinea il profilo delle DH nell'ottica dei sistemi di gestione delle informazioni a 360 gradi. O, in una parola, nel ragionamento sui metodi per un'efficace ed efficiente organizzazione della conoscenza, che trova nei Linked Open Data (da ora LOD), ovvero una possibile implementazione del Web semantico, tanto una metodologia quanto uno strumento. Perché chi si occupa di DH individua nello studio sulle metodologie per la manipolazione delle informazioni, un naturale spazio di ricerca. La nozione di scienza dell'informazione trova una sua puntuale sistematizzazione nella Library and Information Science (comunemente nota come LIS), quindi nel rapporto con il mondo delle biblioteche – metodologie e pratiche connesse – e in generale con i sistemi di classificazione del sapere di origine bibliografica e biblioteconomica. Il contributo originale delle DH alla LIS è proprio quello di usare quegli stessi sistemi finalizzati a organizzare la conoscenza (come ad es. tassonomie, thesauri, ontologie), utilizzando in più il filtro dell'osservazione dell'umanista su dati. Un'osservazione che è interpretazione, perché passa per una prospettiva analitica di impostazione, diremo, filologica, che arricchisce l'osservazione dei dati con la nuova conoscenza dell'interprete (o, diremo meglio, dei diversi interpreti). A queste riflessioni è dedicato il secondo contributo del presente volume, che, con l'analisi del rapporto DH-LIS, vuole delineare l'apporto interpretativo delle DH ai sistemi di gestione dell'informazione, così come nel Web semantico (attraverso i LOD) si vanno configurando.

La lezione appresa dal rapporto fra LIS e DH consente di estendere il ragionamento alla nozione di contesto. Che è il tema del terzo contributo. Se la descrizione uniforme delle fonti, di tradizione biblioteconomica e archivistica, è stata assunta dalle DH come metodologia per la formalizzazione della conoscenza su quelle stesse fonti, allargare lo spettro delle componenti descrittive paratestuali, così come elaborate dagli standard di riferimento, è un vincolo necessario. E la tradizione archivistica, che storicamente ha consolidato la nozione di contesto, consente alle DH di estendere la

prospettiva critica: soggetto produttore, *provenance* e struttura dati (dall'albero al grafo) sono i tre pilastri su cui fondare una nuova metodologia di rappresentazione della conoscenza. Contesto, aggiungeremo, che, secondo la recente riflessione archivistica, è meglio pensare al plurale, ovvero i contesti, le diverse componenti che arricchiscono la descrizione, attraverso quei nuovi filtri di osservazione che il Web semantico sollecita e che determinano quel passaggio dall'albero al grafo che i modelli concettuali impongono.

Ragionare per contesti significa ragionare anche sui diversi livelli attraverso cui la descrizione delle fonti, soprattutto quando calata in un contesto digitale, può essere realizzata, ovvero significa riconoscere nel processo che conduce l'osservatore a restituire la propria interpretazione sull'oggetto osservato, e a documentare questa azione, una serie di step che si riflettono per livelli. Potremmo anche dire che ogni momento del processo di descrizione è un diverso livello attraverso cui osservare l'oggetto nel suo contesto di riferimento. E il contesto varia al variare del livello di osservazione. Questi i ragionamenti che contraddistinguono il quarto contributo del volume.

Il tema successivo, quello del quinto contributo, vuole unire le riflessioni precedenti per usare la nozione di organizzazione della conoscenza di nuovo nell'ottica del Web semantico. O meglio di una sua speciale realizzazione, ovvero i già citati LOD. Nuovo "strumento di corredo" dell'informazione sul Web, o meglio dei dati sul Web, i LOD rappresentano il modo attraverso cui la conoscenza – relativa a risorse bibliografiche, archivistiche o museali – può essere organizzata e diventare uno strumento a uso anche del lettore/fruitore finale. Ripensare ai LOD, e in particolare alle ontologie che stabiliscono i requisiti descrittivi delle entità, come "faccette" e quindi strumenti di filtro per un'esplorazione sapiente di collezioni di dati, trasforma i dataset in ambienti di conoscenza per l'utente, consentendo, attraverso nuovi metodi di visualizzazione, di arricchire l'esperienza del lettore in qualità di fruitore di un sapere potenziale e latente.

Ultimo contributo del volume parte dal tema del lavoro a progetto, che qualifica, fra gli altri, l'approccio delle DH alla realizzazione di contenuti digitali come esito di processi computazionali finalizzati a rispondere a precisi quesiti di ricerca, per inserirlo nel più ampio dibattito sulla gestione del ciclo vitale degli oggetti culturali. Nuove professionalità stanno nascendo sul

Web e per il Web. Riconoscere conoscenze e competenze che i nuovi esperti di gestione dell'informazione digitale, in particolare nel Web semantico, devono avere, è un tema cui le DH non possono esimersi dal governare.

DH e Web semantico sono dunque temi strettamente correlati che, superando la dimensione puramente tecnologica e implementativa, permettono di ripensare alle strategie di valorizzazione della conoscenza proveniente dai dati culturali, per arricchire l'esperienza di chi quei dati può usare per ottenere risposte a domande di ricerca sempre più sofisticate.

Non si vuole in questo volume dar conto di aspetti tecnologici, o di singoli progetti all'avanguardia, se non attraverso alcuni modelli esemplari, ma riflettere su aspetti metodologici, che possano contribuire a valorizzare il ruolo delle DH nel contesto di un Web che attribuisce alla cosiddetta "semantica esplicita" il ruolo di strumento per veicolare l'organizzazione della conoscenza proveniente dai dati culturali.

Lo scopo di questa collezione di saggi è proprio quella di affrontare trasversalmente, o come si dice oggi, attraverso un approccio cross-disciplinare, il tema della conoscenza così come le DH lo stanno affrontando, ovvero assumendo un'ottica speciale sul rapporto storico fra le metodologie computazionali e le discipline umanistiche, in particolare quelle che si occupano del libro e del documento. Diremo che le metodologie di queste discipline si ripensano e si affinano alla luce del Web semantico e le DH contribuiscono a costruire quel ponte di trasmissione che agevola il dialogo fra l'informatica e le discipline umanistiche. Un ponte che, oggi più che mai, richiede ai professionisti che lavorano sui dati e con i dati, di affrontare le sfide di una tecnologia imperante, di farlo in modo consapevole, con un approccio attivo e propositivo nei confronti dei sistemi di trasmissione della nostra tradizione culturale.

Se dovessimo, in una frase, riassumere lo scopo di questo lavoro, diremmo che è nel processo di produzione, conservazione, manipolazione e disseminazione dei dati culturali, ovvero anche nella presa di coscienza dell'ecosistema in cui convivono oggetti eterogeni in sistemi informativi differenti, che le DH assumono il ruolo di meta-disciplina, offrendo alla ricerca, prima di ogni altra cosa, una nuova metodologia trasversale di analisi della conoscenza.

Non lasciamoci allora attrarre dal miraggio della tecnologia fine

a sé stessa, ma ragioniamo sui metodi e sulle domande di ricerca che i sistemi computazionali sono capaci di suggerire agli umanisti, per un nuovo approccio (formale) al sapere. Ed è in questo modo che potremmo cercare di fare chiarezza sul fondamento epistemologico delle DH nel contesto di archivi, biblioteche e musei.

1 Associazione Italiana per l'Informatica Umanistica e la Cultura Digitale (AIUCD), http://www.aiucd.it.
2 European Association for Digital Humanities (EADH), https://eadh.org.
3 Alliance of Digital Humanities Organizations (ADHO), https://adho.org.

1. LA HUMANITIES COMPUTING INCONTRA IL WEB (SEMANTICO)

1.1 Una premessa

DH è un'etichetta che oggi va particolarmente di moda. Ma la sua storia conta ormai un discreto numero di anni di esperienza, di cadute e risalite, di iniziative e sperimentazioni tanto nella ricerca, quanto nella formazione.

Se dovessimo tentare una definizione di DH, ci scontreremmo con una significativa varietà di proposte, che naturalmente originano dai diversi punti di vista attraverso i quali l'incontro fra le scienze umanistiche e il digitale può essere affrontato, descritto e codificato.

La letteratura in materia sta diventando sconfinata,[4] ma senza dubbio la tradizione di questo ambito di studi attribuisce agli studiosi di DH una serie di conoscenze e competenze trasversali, un connubio fra l'esperienza sulle discipline, nel solco delle canoniche arti liberali, e la nozione di computabilità, attribuibile alla macchina.

Computabilità appunto, non solo digitalizzazione. La macchina è in grado di manipolare sapientemente simboli quando adeguatamente istruita e non solo di visualizzare un dato perché espresso in un formato binario.

E in mezzo ci sta la metodologia. La capacità di ripensare al metodo della ricerca delle discipline umanistiche tradizionali in una dimensione formale. Ecco che l'abilità di colei o colui che oggi chiameremmo *digital humanist* è di saper modellizzare l'osservazione di un dominio di interesse in modo tale che la concettualizzazione

che ne deriva consenta alla macchina di restituire conoscenza. Un atto di interpretazione, diremo, basato su procedure, regole e principi che governano un processo che dai dati deve arrivare a restituire sapere. Il risultato è quello di definire una nuova metodologia di ricerca, come esito della riflessione sui metodi e sulle tecniche delle discipline coinvolte: quelle umanistiche e quelle informatiche. Se la ricerca umanistica si ripensa alla luce dei sistemi computazionali, analogamente l'informatica affina i suoi metodi acquisendo l'esperienza degli studi sul testo e sul documento.

Ma partiamo dalle origini. L'Informatica Umanistica,[5] altrimenti chiamata in Italia Umanistica Digitale, si confronta con la tradizione della Humanities Computing[6] europea e statunitense che, dai primi degli anni 2000, ha abbracciato la nozione internazionale di Digital Humanities.[7] Convenzionalmente diremo dal 2004, quando la casa editrice Blackwell ha pubblicato *A Companion to Digital Humanities*,[8] che ancora oggi rappresenta l'avvio di quel passaggio, che non è solo terminologico, da Humanities Computing o anche Humanities Computer Science (secondo la definizione di ACO*HUM[9]) a Digital Humanities.

Fino dalle origini, sulle quali torneremo nella prossima sezione dedicata alla storia della disciplina, la Humanities Computing ha rappresentato un'esigenza nata in un contesto di natura linguistica e letteraria: manipolare stringhe di caratteri, allo scopo di sfruttare la capacità della macchina in termini di rapidità nell'esecuzione delle operazioni e di precisione del risultato. Un aspetto meccanico sì, non ancora metodologico, ma che segnalava l'esigenza di lavorare sui dati testuali, e in particolare sui testi letterari, attraverso sistemi di analisi linguistica puramente quantitativi. Vedremo nella prossima sezione dove e come sia originata questa esigenza, ma per ora ci basti riportare alcune date storiche che, a livello internazionale, ci permettono di riconoscere le origini delle DH attraverso la Humanities Computing: la rivista "Computers and the Humanities" è uscita con il suo primo numero nel 1966,[10] e questo ci mostra quanto, già negli anni '60, il dibattito su temi relativi al rapporto fra discipline umanistiche e sistemi computazionali, necessitasse di documentare quelli che oggi possiamo classificare come i primi esiti della ricerca scientifica disciplinare; negli anni settanta vengono fondate due importanti associazioni: l'europea ALLC, nata nel 1973,[11] e l'americana ACH fondata nel 1978.[12] L'esistenza di due Associazioni, a copertura internazionale, fa capire

quanto la comunità scientifica stesse consolidando la ricerca e sentisse l'esigenza di creare luoghi di scambio e dialogo. I membri di ACH e ALLC sono editor di un'altra importante rivista di settore: "Literary and Linguistic Computing" (fondata nel 1986),[13] e che ancora oggi è uno dei pilastri della divulgazione scientifica della disciplina. Gli anni '80 peraltro, come diremo oltre, sono il vero momento di lancio delle iniziative di settore a livello internazionale. Ma senza dimenticare l'importante ruolo dell'Italia: è del 1962 l'esperienza dell'Almanacco Bompiani dedicato al ruolo del calcolatore in letteratura.[14]

Insomma, linguistica e letteratura, assieme al ruolo della macchina di calcolo come esecutore di algoritmi, sono il nostro punto di partenza, necessario per tracciare una cronistoria della disciplina e comprendere come il Web abbia poi modificato in modo significativo, a partire dagli anni 2000, la ricerca di settore. L'approccio che si intende adottare nelle prossime pagine, non vuole tanto documentare una prospettiva strettamente disciplinare (metodi computazionali applicati alle singole discipline umanistiche), ma vuole essere una chiave di lettura trasversale (metodologie per l'acquisizione di conoscenza attraverso processi computazionali), pur riconoscendo le esigenze specifiche della ricerca nei diversi settori disciplinari, o, meglio, pur nella consapevolezza della diversità dei metodi computazionali da adottare a seconda delle esigenze specifiche delle domande di ricerca, ma sempre con l'obiettivo dell'acquisizione (trasversale rispetto alle discipline) di nuovo sapere.

Avviamo la nostra riflessione recuperando il modello ACO*HUM sopra menzionato per dire che cosa questa disciplina, fino dalle origini e ancora oggi, non è: non è mero uso di risorse digitali (consultazione di dati accessibili in un qualche formato elettronico), non è digitalizzazione base di fonti (acquisizione di dati attraverso sistemi di scansione), non è alfabetizzazione all'uso del calcolatore (*literacy*), non è costruzione di siti Web (puro output di formato), non è *data entry* (immissione di dati in un'applicazione), pure in contesto digitale. Ma è la capacità di rappresentare criticamente le fonti, e l'informazione complessa e multiforme da queste veicolata, attraverso metodi formali, ovvero è l'abilità di usare regole, schemi, strutture e modelli per fare dei dati, naturale espressione del contenuto degli artefatti culturali, risorse computabili.

Per iniziare a ragionare sulla Humanities Computing diremo, con

Manfred Thaller, che sono le strutture di dati e gli algoritmi a guidare l'applicazione di sistemi computazionali alle discipline umanistiche, ovvero a governare la definizione dei metodi che vanno dalla rappresentazione dell'informazione (*data structures*) all'ottenimento di risposte a specifici quesiti di ricerca (*algorithms*):

> [...] we will attempt to define the core of all applied computer sciences in terms of the traditional combination of data structures and algorithms, applied to the requirements of a discipline:
> • The methods needed to represent the information within a specific domain of knowledge in such a way that this information can be processed by computational systems result in the data structures required by a specific discipline.
> • The methods needed to formulate the research questions and specific procedures of a given domain of knowledge in such a way as to benefit from the application of computational processing result in the algorithms applicable to a given discipline.[15]

La riflessione di Thaller è molto interessante per il punto di vista sulle DH che si vuole qui abbracciare, perché focalizza su alcuni concetti chiave, che rappresentano il cuore dell'origine della Humanities Computing: la nozione di metodo, come lo strumento necessario, prima di qualunque altro, per stabilire una procedura – che diventa la metodologia come cuore delle DH; il concetto di rappresentazione dell'informazione come base dell'impiego di sistemi computazionali – che implica che il tipo di rappresentazione digitale scelta veicola e stabilisce le possibilità della computazione; il ruolo del dominio di conoscenza cui applicare il metodo, e quindi il ruolo delle discipline nel definire il più adeguato intervento computazionale – che richiede consapevolezza dei problemi di ricerca delle scienze umane; l'importanza delle domande di ricerca, che guidano la progettazione e quindi l'implementazione – che richiama il tema del lavoro orientato alla realizzazione di un progetto; e, per concludere, l'algoritmo, come risultato della riflessione sul metodo computazionale – che rappresenta il cuore dell'applicazione. Temi su cui torneremo lungo il nostro percorso.

Per chiudere questa premessa, un'ultima considerazione sul rapporto DH e CH viene dalla classificazione delle attività proprie di questi due ambiti, presentata da Dino Buzzetti ad AIUCD 2019.[16] Da questa slide (Fig. 1) cogliamo lo spessore teorico (e di conseguenza pratico e applicativo) di un rapporto che cercheremo di argomentare nelle prossime pagine. Digitale da un lato, computazionale dall'altro. In due frasi: 1) DH = digitale, come

strumento di visualizzazione in ottica di creazione di strumenti di interfaccia per oggetti multimediali destinati all'utente finale; 2) HC = computazionale, come formalizzazione di un'argomentazione logica che documenta l'interpretazione nei termini di un metodo per la manipolazione dei dati utile all'acquisizione di nuova conoscenza. Quello che cercheremo di raccontare è come il Semantic Web consenta di sposare queste due prospettive (DH + HC) in una nuova visione dell'interazione uomo-macchina in cui è l'organizzazione della conoscenza il filo conduttore.

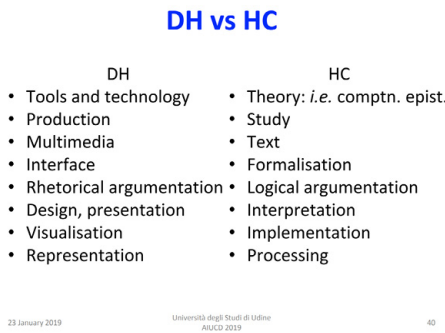

Figura 1 - *DH vs HC* secondo la proposta di Dino Buzzetti, in *AIUCD2019*, slides, http://aiucd2019.uniud.it/wp-content/uploads/2019/02/buzzetti-udine_2019.pdf, slide 40

1.2 Cronistoria

La storia della disciplina, che intendiamo delineare, risale, come ormai noto ai più, agli esperimenti di linguistica computazionale condotti in Italia, sul finire degli anni '40, dal gesuita Padre Roberto Busa17 che, schede perforate alla mano, si dedica all'indicizzazione dell'intero corpus di San Tommaso d'Aquino.18 Operazione utile a velocizzare azioni altrimenti lunghe, nonché foriere di un margine significativo di errore, se condotte a mano. Ma pur sempre un primo esperimento di uso della capacità computazionale della macchina, intesa come manipolatore di stringhe, per restituire un risultato esito delle procedure canoniche di analisi di un corpus testuale – in particolare per la realizzazione di concordanze. Leggendo il lavoro di Busa oggi, diremo che esso rappresenta, insieme, le origini della Humanities Computing e le nuove tendenze correnti nella gestione

di processi di analisi del testo nel sistema dei big data,[19] dove si adotta l'analisi statistica di grandi quantità di dati in un'ottica che chiameremo del *distant reading*,[20] ovvero attraverso utilizzo di metodi computazionali nella gestione dei dati letterari, finalizzati alla scoperta di qualcosa di nuovo.

Dal punto di vista dei sistemi computazionali, in quegli stessi anni in cui Busa opera, il ragionamento logico è affidato agli esiti del lavoro di Alan Turing, e in particolare alla concretizzazione della teorizzazione della nozione di algoritmo attraverso la nota Macchina di Turing.[21] La modalità con cui la macchina esegue azioni è la traduzione di un insieme di step espressi in modo formale che conducono alla restituzione di un esito. Partendo da A, si può optare (*if*) per percorrere la strada A1 o A2, e la conseguenza di questa decisione porta ulteriori ramificazioni del percorso per arrivare alla fine del processo. Ed è innegabile che questa modalità di approccio alla conoscenza obblighi a esplicitare un punto di vista, a usare delle regole ben precise per rappresentare quel punto di vista, e a scegliere un modello, che in questo caso è uno schema di traduzione di quel punto di vista.

Ecco allora che, di nuovo, voler abbracciare un'impostazione computazionale significa osservare la realtà e scegliere le strutture più adeguate alla sua formalizzazione, al fine di acquisire conoscenza attraverso l'opzione di un processo. Le strutture di dati hanno sempre ricoperto un ruolo importante nella storia delle DH. Lista, albero, grafo e tabella rappresentano le astrazioni attraverso cui interpretare l'osservazione dei dati. E la scelta della struttura stabilisce la serie delle operazioni (concrete) che sui dati possono essere poi condotte.

È allora importante ricordare che è sempre di quegli anni il famoso Memex di Vannevar Bush, ingegnere statunitense che, con la sua *Memory Extension*, o anche con le riflessioni sui principi del ragionamento associativo,[22] è considerato il teorizzatore della nozione di ipertesto; ipertesto che, a sua volta, a voler rapidamente documentare le origini, è termine coniato nel 1965 da Ted Nelson.[23] L'origine della riflessione di Bush nel suo *As we may think*, rappresenta proprio l'esigenza di superare processi strettamente tassonomici di osservazione del reale, per sposare la naturale interconnessione fra livelli appartenenti a rami diversi di una gerarchia. Che è il modo attraverso cui la mente umana ragiona: ovvero per associazioni. Ed è la modalità associativa che da una

struttura gerarchica, ovvero l'albero, permette di passare a una diversa struttura cioè quella a rete, o anche a grafo. Non è un caso se l'ipertestualità (lettura potenzialmente multisequenziale e multilineare fra unità informative interconnesse) basa su questa nozione la teoria che sta a fondamento del funzionamento del Web come oggi lo conosciamo (e su cui torneremo).

Una breve cronistoria delle DH non può non considerare però lo stretto legame che le teorie, le metodologie e le tecniche computazionali hanno avuto non solo con i principi a fondamento dell'analisi del testo (tecniche per elaborare automaticamente indici, frequenze e concordanze), ma anche con il mondo degli archivi e delle biblioteche.

Siamo negli anni '60 e, mentre viene coniato il termine *hypertext*, iniziano a prendere forma i primi esperimenti americani di creazione dei cataloghi elettronici delle biblioteche, ovvero dei primi modelli di OPAC.[24] OPAC che condividono, non a caso, gli anni '60 non solo con l'ideazione, da parte della Library of Congress, del formato MARC[25] per la rappresentazione dell'informazione bibliografica, ma soprattutto con le prime sperimentazioni di quello che sarà chiamato database, termine coniato per indicare sistemi nuovi di organizzazione di collezioni di dati. E torniamo di nuovo all'importanza del modello di rappresentazione, ovvero quello gerarchico e a rete prima (anni '60) e relazionale poi (con Codd, dagli anni '70).[26] Modello, quello relazionale, divenuto poi fra i più utilizzati, in particolare perché trova nella struttura dati tabellare la sua naturale forma espressiva. Di nuovo, è la tipologia dei dati osservati, assieme alle possibilità di ottenere da quei dati un'informazione, che determina la scelta di un modello.

Non andrà dimenticato che nel 1968, sul "Journal of the Association for Information Science and Technology" (JASIST), Borko dà una definizione di *information science* a oggi (per lo più) condivisa dalla comunità scientifica (si veda par. 2.1).[27] Si tratta di quell'approccio all'informazione che sta a fondamento della nascita delle prime *information schools* (meglio note come I-school), ovvero di quello che sarà l'embrione della LIS.[28]

E le scienze dell'informazione sono strettamente connesse alle scienze della documentazione. Ma dobbiamo risalire ben addietro per cogliere le origini di queste riflessioni, in particolare addentrarci nella teoria dei sistemi di classificazione: dalle proposte

di organizzazione della conoscenza fine ottocentesche di Melvil Dewey (prima redazione della sua *Dewey Decimal Classification* DDC, 1876[29]) e dell'antagonista Henry Bliss (primo articolo sul tema 1910[30]), passando per gli studi sulla documentazione di Paul Otlet[31] e Henri La Fontaine degli anni '30 del secolo scorso (dal Mundaneum[32] alla pubblicazione della *Universal Decimal Classification* UDC, dal 1905[33]), assieme alle proposta di classificazione sintetico-analitica di Shiyali Ramamrita Ranganathan con la *Colon Classification*, altrimenti nota come "classificazione a faccette" (CC, dal 1933[34]); insomma, il problema di rappresentare e gestire l'informazione trova nel contesto della ricerca condotta nelle biblioteche il sostrato teorico per i modelli computazionali degli anni '60.[35]

Se il rapporto fra biblioteche e tecnologie inizia ben prima del calcolatore a uso personale, altrettanto precoci sono le relazioni fra i metodi computazionali e l'ecdotica. Degli anni '70 sono infatti le prime sperimentazioni in campo filologico, orientate a risolvere il problema della collazione automatica e ad acquisire l'orizzonte stemmatico, in una prospettiva di automatizzazione dell'albero genealogico delle dipendenze fra i codici di una tradizione testuale. Sono anni segnati dall'operato di Jacques Froger, anche nella critica al modello elaborato da Henri Quentin,[36] e sintetizzate da Jean Irigoin e Gian Piero Zarri.[37] Un primo collegamento, anche se di natura ancora strettamente meccanica, fra la filologia e la manipolazione computazionale delle fonti testuali,[38] che è un elemento importante per l'evoluzione delle DH (anzi diremo della Humanities Computing, come era ancora chiamata allora). Ecco che ancora la testualità è al centro degli interessi di ricerca (basti sfogliare gli indici delle già citate riviste "Computing in the Humanities" e "Literary and Linguistic Computing" per cogliere le istanze della ricerca scientifica delle origini) e in particolare la possibilità di esplorare una testualità complessa per rispondere a domande di ricerca altrettanto complesse, ma da un certo punto di vista semplici se delegate a una macchina (semplici non si riveleranno poi, quando le domande di ricerca si faranno computazionalmente più sofisticate o quando si chiederà alla macchina di valutare il "peso" di una variante[39]).

Gli anni '70 sono anche quelli in cui la nozione di catalogo (la descrizione a livello di paratesto) incontra il testo pieno delle opere; ovvero è l'epoca della formazione del corpus dei testi greci TLG,[40]

dell'OTA[41] e, dagli anni '80, della collezione di testi latini del PHI,[42] quindi delle prime raccolte full-text in *Machine Readable Form*, su cui effettuare interrogazioni a scopo di analisi (quindi ancora indici, frequenze e concordanze). Ma, attraverso queste prime collezioni strutturate di testi pieni accessibili alla comunità, sono anche gli anni di una prima definizione embrionale del concetto di biblioteca digitale.[43]

Dagli anni '40 arriviamo, attraverso questi significativi ed essenziali momenti storici, agli anni '80, caratterizzati dalla necessità di un consolidamento. Questo è il periodo segnato tanto dall'elaborazione di software capaci di affrontare le sfide della ricerca di settore, quanto dalla definizione di standard internazionali che intendono superare frammentarietà e divisioni, per favorire dialogo e interscambio. In quarant'anni di esperienza sul rapporto fra computing e humanities, ma anche con la nascita dei primi software applicativi, l'esigenza di standardizzare i processi, in ottica di interoperabilità, diventa cioè una necessità. E l'attenzione si sposta dall'automazione, alla condivisione di principi, regole e modelli, prima di tutti teorici e metodologici.

Sono infatti degli anni '80 la riflessione sui linguaggi di markup dichiarativo e la nascita dell'SGML (1986),[44] assieme alle prime iniziative della TEI (1987):[45] sintassi e grammatica sono cioè lo strumento per l'interoperabilità, tanto sul piano sintattico quanto su quello del vocabolario di interscambio. In particolare, l'esigenza, già da allora, di svincolare i dati dal software capace di leggerli, porta alle prime sperimentazioni sui metalinguaggi, che trovano, proprio nella sintassi dell'SGML, la più compiuta forma di traduzione di questa istanza. Con alcuni concetti chiave alla base di un nuovo approccio alla testualità: 1) l'adozione del modello ad albero, ovvero la necessità di rappresentare, attraverso un sistema gerarchico, gli elementi costitutivi dell'unità informativa – che sia il testo, ovvero la sequenza di stringhe di caratteri, e/o il documento, ovvero il supporto materiale che veicola il testo; 2) la necessità di esprimere questi elementi, ovvero le componenti della marcatura, attraverso un vocabolario descrittivo condiviso per una classe di documenti affini. È in questo contesto che la TEI, nel dominio di quella classe rappresentata dal testo, diremo, umanistico, inizia a muovere i primi passi, che condurranno all'elaborazione di quelle *Guidelines*,[46] ancora oggi strumento di riferimento. Con la TEI, divenuta standard per la comunità, inizia ad affermarsi l'esigenza di

stabilire il modello di codifica più adeguato a esprimere l'interpretazione dell'ermeneuta sul rapporto testo/documento. Con tutte le complessità che il dibattito critico ha sollevato.[47] Con TEI si afferma cioè il principio della modellazione intesa come assunzione di un punto di vista sull'oggetto d'analisi. L'organizzazione dello Schema per "classi" si basa tanto sulla classificazione delle tipologie testuali e documentarie (prosa, poesia, testo drammatico, ma anche lettere, dizionari, parlato, manoscritto ecc.), quanto sui bisogni diversi della comunità, determinati dagli obiettivi che con la marcatura si possono esaudire (edizione di un testo, nelle sue diverse modalità, trascrizione di una fonte primaria, descrizione di un manoscritto, analisi linguistica, narratologica, retorica ecc.).

Ed è proprio negli standard che anche il dominio dei beni culturali consolida, nel corso degli anni '90, le proprie esigenze. Standard metodologici, capaci di stabilire auspicabili linee guida condivise, necessarie a favorire l'interscambio, ma anche primi software di inventariazione per gli archivi[48] assieme al consolidamento del catalogo collettivo SBN, nato già nel decennio precedente.[49] Diciamo che la definizione dei primi software va di pari passo con l'avanzamento delle riflessioni sul metodo. Più il metodo si affina (e con il metodo i modelli della descrizione), più l'applicativo è in grado di rispondere ai bisogni di bibliotecari e archivisti.

Ecco che gli anni '90 determinano, nel contesto degli archivi, la nascita dell'ISAD (prima edizione 1994),[50] il pilastro della descrizione delle fonti documentali, che sarà alla base della nascita dello standard di struttura EAD (del 1998),[51] una DTD a base SGML prima e XML poi. Ma sono anche gli anni dell'elaborazione di FRBR (prima edizione 1998),[52] che ha occupato tanta parte della riflessione sulla necessità di elaborare modelli per l'analisi dei contenuti trasmessi dalle fonti materiali, non solo di natura bibliografica (per cui si veda cap. 4).

Il metalinguaggio XML (sempre 1998),[53] in particolare, ha svolto un ruolo importante nel contesto della Humanities Computing. È stato il primo tentativo di estendere le capacità di un linguaggio di marcatura, di renderlo universale, di usare una sintassi di interscambio, lasciando massima libertà nella scelta del vocabolario di marcatura. Ed è in questo contesto che DTD come TEI e EAD hanno iniziato a muovere i primi passi. Ma non solo. È la DTD, a base SGML prima e XML dopo, ovvero HTML (prima versione

pubblica del 1993), quella che ha contribuito a una diffusione massiccia di contenuti informativi, anche di natura umanistica, ovvero anche a riconoscere, come vedremo a breve, l'importanza della presenza di un utente finale cui destinare gli esisti degli approcci computazionali, fino ad allora riservati a una ricerca scientifica fondamentalmente auto-referenziale.

Ma prima di arrivare alla rivoluzione del Web, determinata anche dal ruolo dello Schema HTML, qualche parola va spesa sul ruolo dei metalinguaggi di markup e delle grammatiche associate. Se una sintassi condivisa, ovvero quella del metalinguaggio, ha contribuito a svincolare l'oggetto digitale prodotto dall'applicazione necessaria a produrlo, le grammatiche hanno fornito la giusta soluzione all'esigenza di standardizzare il vocabolario. E in fondo avere un vocabolario comune per descrivere le cose, siano esse il concetto di soggetto produttore o di attribuzione di responsabilità o anche genericamente di nome di persona o ancora di data o luogo, è la strada per la condivisione. E allora edizioni scientifiche digitali di testi, record catalografici, strumenti di corredo archivistici si sono potuti giovare di una lingua franca, o meglio diremo di lingue franche disciplinari e specializzate, attraverso cui rendere possibile la formalizzazione del percorso interpretativo, anche in dimensione partecipata, perché condivisa.

Scrive John Unsworth, alla luce del passaggio da Humanities Computing a Digital Humanities (proprio nel fatidico anno 2004):

> we are, I think, on the verge of what seems to me the third major phase in humanities computing, which has moved from tools [1.] in the 50s, 60s, and 70s, to primary sources in the 80s and 90s [2.], and now seems to be moving back to tools... [3.]. I think we are arriving at a moment when the form of the attention that we pay to primary source materials is shifting **from digitizing to analyzing, from artifacts to aggregates, and from representation to abstraction**.[54]

Ci muoviamo dunque dalla digitalizzazione all'analisi, dagli artefatti agli aggregati, dalla rappresentazione all'astrazione. Temi importanti per riflettere sul ruolo che ora il Web sta esercitando, tanto sui sistemi di rappresentazione, quanto di manipolazione, di dati, informazione e conoscenza. Temi anche necessari per focalizzare su alcuni concetti chiave del processo di gestione degli oggetti digitali: analisi, aggregazione e modellazione. Ovvero, parafrasando Unsworth, diremo che: 1) non basta più acquisire digitalmente un oggetto analogico, ma è necessario che la capacità

computazionale della macchina come calcolatore, sia usata nell'analizzare l'oggetto digitale prodotto; 2) ogni oggetto non è un atomo isolato, ma un'entità in dialogo con altre entità correlate a costruire nuovi aggregati di conoscenza; 3) i procedimenti astrattivi sono fondamentali per modellare la realtà osservata e produrre una descrizione formalizzata (computabile) di quella realtà.

E vediamo come, arrivati agli anni '90 del secolo scorso, il Web ha affrontato questi temi.

1.3 La rivoluzione del Web

L'avvento del Web è senz'altro il fenomeno che più ha rivoluzionato le pratiche dell'informatica umanistica. Pur ideato, assieme alla sua lingua franca HTML, all'inizio degli anni '90, è solo alla fine del millennio che il Web ha profondamente segnato le sorti della Humanities Computing. Meglio ha determinato quel passaggio, di cui prima si diceva, dalla dimensione della Humanities Computing a quella delle Digital Humanities.

Come abbiamo detto, certamente la Humanities Computing è strettamente legata alla sfera letteraria e linguistica, e l'approccio alla computabilità nelle discipline umanistiche delle origini è, in generale, fortemente orientato alla specificità dei metodi formali per ogni singola disciplina umanistica. I metodi per computare i dati (testi e documenti) si attestano su un livello *machine-oriented*, una dimensione in cui l'utente finale non è la priorità del risultato della formalizzazione; formalizzazione che è invece finalizzata a risolvere i bisogni di ricerca dello studioso che, attraverso metodi automatici, acquisisce conoscenza dai dati, primariamente di natura testuale.

È con il Web che si inizia a ragionare in un'ottica trasversale, ovvero: 1) sulle metodologie che tutte le discipline umanistiche, non solo testo-centriche, condividono (markup, basi di dati, ipertesti, sistemi di classificazione e modellazione), 2) sulla condivisione dei processi (rappresentazione, analisi, interrogazione e accesso) e 3) sull'importanza del ciclo vitale dei documenti digitali, così come lo si può dedurre dalla serie delle attività che lo governano (produzione, descrizione, gestione, conservazione, disseminazione, uso e soprattutto riuso).[55] L'accesso distribuito alle risorse diventa un aspetto fondamentale e l'utente è collocato nel

cuore del sistema di fruizione. La pubblicazione sul Web degli oggetti digitali è il naturale output della ricerca di settore e l'approccio orientato ai dati (*data-driven*), al contempo esito e conseguenza delle istanze della *Web analytics*, è centrale per fornire servizi sempre più sofisticati al fruitore. L'utenza, nella sua dimensione "profilata", diventa il centro del processo comunicativo (e decisionale). Non sarà un caso che il già citato *A Companion to Digital Humanities*, edizione 2004, veda l'approccio canonico della Humanities Computing sposarsi con la nuova apertura disciplinare delle DH e quindi: 1) parli di storia del rapporto fra singole discipline umanistiche e tecnologie (dagli studi classici alla musicologia, dall'archeologia alla storia dell'arte, dalla lessicografia alla linguistica, dagli studi letterari alle arti performative); 2) dia contestualmente valore ai principi condivisi dalle discipline stesse (metodi e modelli) e alle applicazioni che ne derivano (progetti e ambiti), e solo in un'ultima parte ragioni sulle azioni (sarà sufficiente leggere l'indice, dove le sezioni sono: *history; principles; applications; production, dissemination, archiving*). Allo stesso tempo, non si può non rilevare come l'edizione del *Companion* del 2016 abbandoni la Humanities Computing disciplinare per ragionare in termini di azioni, ovvero di valorizzazione dei processi, dando contestualmente all'utente un ruolo importante di fruitore delle risorse prodotte (le quattro sezioni in indice sono: *infrastructures, creation, analysis, dissemination*).

Il Web insomma ha segnato le sorti delle discipline umanistiche. Sistemi informativi archivistici, biblioteche elettroniche e poi digitali, edizioni scientifiche di testi e documenti, si sono giovati, come si diceva, di questa innovazione. O meglio, hanno saputo trovare nel Web un luogo non tanto (quantomeno alle origini) per la manipolazione di contenuti in una prospettiva computazionale, ma soprattutto un luogo, disciplinarmente agnostico, per la disseminazione e la pubblicazione dei risultati della ricerca.

Il Web stesso, nel suo percorso di evoluzione, è cambiato: da editor per ipertesti, nel progetto originario del suo ideatore,[56] è diventato ambiente di fruizione riservato agli specialisti del codice (Web 1.0); da sistema di accesso passivo ed elitario, il Web si è ripensato in termini di ambiente di produzione collettiva e partecipata del sapere (Web 2.0); da collettore di informazione sotto forma di documenti ipertestuali, il Web è diventato il luogo per ospitare dati semantici (Web 3.0).

L'evoluzione del Web ha insomma determinato un cambiamento importante nei sistemi di rappresentazione digitale del contenuto trasmesso dalle risorse informative. Questo percorso ha profondamente influenzato tanto il lavoro degli istituti culturali nella gestione del sapere custodito nei cataloghi delle biblioteche e negli inventari archivistici, quanto nel mondo della ricerca, che ha iniziato a utilizzare fonti primarie (librarie e documentali) per produrre edizioni scientifiche digitali o raccolte di oggetti culturali, come esito della valorizzazione di quelle fonti, da rendere fruibili attraverso una piattaforma aperta e distribuita, ovvero il WWW.

Diremo che, già nella fase del Web 1.0, gli archivi hanno iniziato a pubblicare su Web gli strumenti di corredo realizzati attraverso applicativi software proprietari, ma anche a disseminare le fonti digitali, acquisite dagli esemplari analogici;[57] le biblioteche, che già molto prima del Web rendevano disponibili i loro cataloghi attraverso terminali dedicati o con collegamento Telnet, hanno pubblicato i loro OPAC su Web, mettendo a disposizione dell'utente finale un sistema di query con interfacce più amichevoli;[58] sono comparsi i primi ipertesti letterari,[59] o meglio la teoria sugli ipertesti ha trovato nel Web il naturale luogo di valorizzazione e sono comparse le prime edizioni non legate a software particolari,[60] ma in grado di utilizzare i linguaggi di markup dichiarativi (SGML/TEI e poi XML/TEI) per consegnare all'HTML i risultati della *restitutio textus*, documentando la storia della tradizione.[61] Ancor più delle edizioni critiche, il Web è stato il luogo utile a ospitare numerose edizioni diplomatiche o interpretative – letterarie,[62] ma anche documentali[63] –, capaci di tradurre in linguaggio formale tutti i fenomeni della tradizione, in particolare manoscritta.[64]

Ma gli stessi linguaggi del Web si sono ripensati, in un'ottica di rinnovamento delle metodologie della rappresentazione, e l'inizio degli anni 2000 ha segnato una nuova fase. Come si anticipava, dall'attenzione al documento si è passati alla formalizzazione del dato, ovvero alla traduzione degli elementi del documento in composti atomici, identificati e qualificati, in grado di veicolare l'informazione, e capaci di restituire conoscenza solo quando arricchiti dalle relazioni. Lo scopo finale di questa trasformazione è quello di acquisire conoscenza dagli oggetti digitali attraverso i dati, veicolati da quegli oggetti, in relazione (potenzialmente multipla e trasversale). E all'HTML si sono affiancati linguaggi nuovi: si è standardizzato il set di caratteri attraverso Unicode, si è

confermato il ruolo di XML nei sistemi di descrizione e interscambio, si è valorizzato il ricorso ai *namespaces* e agli Schemi (ad es. TEI) come strumento di disambiguazione, si è sostituto lo stabile URI al volatile URL, si è iniziato a utilizzare un sistema di descrizione dei dati che ragiona per triple, ovvero RDF, sono state introdotte le ontologie anche nel Web. Insomma, si è avviato il Web semantico, il cui fondamento teorico è che il dato va analizzato e studiato con la consapevolezza delle sue molteplici relazioni, che sono tanto eterogenee quanto ramificate, e vanno elaborati modelli concettuali, ovvero approcci astrattivi dell'osservazione dei dati, nella forma di ontologie, o anche di descrizioni formalizzate.

1.4 La semantica (del Web) e le humanities

Da quando nel 2001 Berners Lee ha presentato il progetto del Web semantico (altrimenti noto come Web 3.0)[65] possiamo dire che molta strada è stata fatta. Applicazioni, progettisti ma anche utenti del Web hanno saputo ricollocare il ruolo dell'informazione digitale alla luce dell'arricchimento concettuale che il nuovo modello di rappresentazione ha in qualche modo imposto. Se la definizione di nuovi standard, ma anche il recupero di linguaggi consolidati, sono stati un elemento importante nell'elaborazione del *Semantic Web stack*,[66] cioè nella sua architettura per *layers*, una consistente novità è stata la ridefinizione stessa del concetto di informazione, assieme alla necessità di configurare il ruolo dei LOD nello spazio di un Web ripensato.

Un elemento interessante dello *stack*[67] è che la realizzazione del Web semantico passa tanto per la definizione della struttura logica da conferire al Web, basata sui principi di concettualizzazione e astrazione (*Concept and Abstraction*), quanto per la scelta degli strumenti, basata su specifiche tecniche e soluzioni tecnologiche (*Specification and Solutions*). Ecco quindi che le nozioni di piattaforma Web (*web platform*), di sintassi (*syntax*), di struttura della rappresentazione della conoscenza (*knowledge representation structure*) e ancora di regole (*rules*) e di semantica (*semantics*) – quindi riflessioni che riguardano la concettualizzazione in un procedimento di astrazione –, si sposano con URI/IRI, HTTP, Unicode, XML, Turtle, RDFa, RDF, OWL, RDFS e SKOS che sono invece linguaggi, protocolli e in generale standard che si riferiscono

al piano dell'implementazione e sono quindi sì scelte tecnologiche, ma esito di un approccio che prima di tutto è concettuale. Concettuale, perché orientato a risolvere problemi di: identificazione univoca di risorse (URI/IRI), protocolli di trasmissione (HTTP), codifica dei caratteri (Unicode), formati di interscambio (XML), rappresentazione della conoscenza basata sul principio di un asserto tripartito (RDF), assieme alle sue serializzazioni (Turtle, RDFa, Microformats),[68] ontologie e Schemi ovvero modelli concettuali (OWL, RDFS, SKOS).[69]

Il Web semantico allora è prima di tutto un nuovo modo di intendere la strutturazione dei dati, che copre il terzo e il quarto livello del nostro *stack*, in cui: 1) RDF ci insegna a ragionare per triple, ovvero asserti dotati di un soggetto, un predicato e un oggetto; 2) con RDFs aggiungiamo un vocabolario, che ci permette di descrivere tanto le risorse quanto le relazioni fra risorse e disponiamo di costruttori per la descrizione dei tipi di oggetti (le classi), dei tipi di gerarchie (sottoclassi) e delle proprietà, che rappresentano le caratteristiche base degli oggetti; 3) con OWL la semantica del linguaggio si arricchisce ulteriormente, con nuovi costruttori, regole e assiomi nel solco del ragionamento logico.

Sulla nozione di semantica, vale la pena una brevissima riflessione. Non sarà un caso che lo stesso linguaggio per la creazione di pagine Web, ovvero l'HTML, nella sua versione 5.0, rilasciata ufficialmente dal W3C nel 2014, abbia introdotto la nozione di semantica per riferirsi a quei nuovi elementi in grado di rappresentare l'architettura informativa a livello di interfaccia (elementi logici della pagina: *header*, *footer*, *article*, *aside*, *section*) e stabilire un vocabolario comune per attribuire un ruolo alle gabbie logiche dello specchio della pagina.[70] Vera semantica non è, siamo solo nel tema della standardizzazione del dizionario per l'annotazione, ma lo sforzo è continuato con il già menzionato RDFa, ovvero con l'inserimento, direttamente dentro al codice HTML, di asserti in forma di triple, ovvero con l'*embedding* di una allusione a una semantica esterna (classi e proprietà) in una struttura sintatica (e su RDFa torneremo).

Ma ripensiamo ai dati. Le scienze umane non sono rimaste immuni da questa trasformazione del Web. E le fome della rappresentazione del testo digitale si sono dovute confrontare con una serie di linguaggi formali, che sono anche però nuovi modi di intendere l'informazione. Come abbiamo già detto, non è più il solo

documento, nella sua dimensione globale di risorsa informativa, che trasmette l'informazione, ma è invece il dato l'oggetto su cui spostare l'attenzione. Questo ha significato fare uno sforzo per tradurre i documenti in dati, ovvero anche per enucleare le entità capaci di qualificare il contenuto del documento e metterle in relazione per restituire un significato.

In particolare, nel contesto del Web semantico, il fenomeno dei già citati LOD[71] ha contribuito a realizzare il movimento del *Web of data*.[72] L'approccio cosìddetto "data centrico" ha sottratto l'ambiguità e l'indeterminatezza di sequenze di stringhe di caratteri non interpretate, tipiche del Web 1.0, con un sistema di rappresentazione formale in grado di consentire alla macchina di comprendere il dato, trasformandolo il conoscenza.

Detto altrimenti, da un approccio tipicamente documento-centrico del Web 1.0, passando per una fase ibrida del Web 2.0,[73] ciò che distingue il Web 3.0 è la scelta di un approccio data-centrico al concetto stesso di informazione. Come abbiamo anticipato, il Web 1.0 ha individuato nel documento HTML il focus della rappresentazione digitale e nel link @href lo strumento per creare collegamenti fra documenti. Il Web 2.0 ha introdotto l'importante concetto della (presunta e ipotetica) democratizzazione del sapere, concentrando l'attenzione sull'idea di un ambiente collaborativo e partecipato, in cui il *crowdsourcing*, o anche la generazione di contenuti da parte di un utente non più passivo fruitore di pagine Web, ma attivo produttore di nuovo sapere, è la chiave per la valorizzazione. Il Web 3.0 ha spostato l'enfasi sul dato atomico, la risorsa qualificata attraverso un frammento identificato univocamente, e sulle relazioni tipizzate, ovvero semantiche.

La semantica, che nel Web 3.0 è rappresentata dalla modellazione concettuale, base per la costruzione di ontologie fondazionali e di dominio – ovvero di strumenti per l'identificazione di classi e proprietà in grado di astrarre le *features* di una realtà osservata e consentire il ragionamento logico e inferenziale (regole e assiomi) –, diventa, con i LOD, un modo per attribuire ai vocabolari la funzione di connettori fra le entità e la loro tipizzazione logica (ad es. un URI riferito a "Francesca Tomasi" connette l'entità a una classe di tipo "persona"; un URI "Francesca Tomasi" e un URI che identifica il "presente volume" sono collegati dalla proprietà "autore", che specifica il ruolo di quella persona).

Se dunque, da un lato, il Web ha progressivamente costretto a

spostare l'attenzione dalla manipolazione alla disseminazione, dal computing al digital, dall'altro esso ha segnato l'idea della semantizzazione dei contenuti culturali in un nuovo contesto di rappresentazione di quelle che potremmo definire "entità concettuali". Tornando alle nostre strutture dati, con i LOD si vuole costruire quello che oggi possiamo chiamare un *knowledge graph*, o anche grafo della conoscenza, ovvero interconnessioni semantiche fra entità disseminate. E questo processo di trasformazione sta avvenendo anche nel contesto del patrimonio culturale, dove sta nescendo il *cultural heritage knowledge graph*, che emerge all'interno di un LOD *cloud* dei beni culturali, in cui le istituzioni, che conservano dati relativi a eventi, luoghi, persone e concetti, cercano sempre più di dialogare in un sistema di relazioni multi-direzionale.[74] Come vedremo oltre, archivi, biblioteche e musei hanno iniziato a pensare ai LOD come a quello strumento che, se da un lato consente di riportare i dati in un ambiente aperto (open data), dall'altro è l'occasione per costruire una rete del MAB[75] o anche del LAM/(G)LAM[76] che valorizzi il patrimonio in una prospettiva di dialogo e reciproco arricchimento (linked).[77] E anche le edizioni di testi, in particolare, hanno saputo trarre dal Web semantico, e dal modello a grafo, nuova ispirazione, ovvero i ricercatori si sono trovati a ripensare il concetto di edizione, tanto letteraria quando documentale, nei termini di una nuova esperienza di rappresentazione digitale della testualità.[78]

Quando nel 2012 Google ha avviato il suo personale progetto di *knowledge graph*,[79] allo scopo di restituire all'utente una *infobox* in grado di pescare da diverse risorse Web, informazioni relative a oggetti, persone, luoghi e organizzazioni (Fig. 2), di presentarle come dati strutturati e di proporre una serie di potenziali relazioni, ovvero collegamenti fra dati, è emersa un'esigenza, che nei prossimi capitoli si cercherà di raccontare. Mettere a disposizione dell'utente del Web semantico dati affidabili, documentati, scientifici, in grado di arricchire autenticamente l'esperienza di navigazione sul Web attraverso l'accesso a informazioni supportate dall'indispensabile corredo paratestuale, metatestuale e intertestuale; ovvero, diremo, ciò che oggi l'utente esperto vuole è l'accesso a quella conoscenza che Google non è in grado di restituire con progetti, pur eccellenti, come Dbpedia[80] o Wikidata,[81] in cui il portato del patrimonio culturale è solo parzialmente valorizzato rispetto al potenziale che potrebbe derivare dalla mole di dati (e metadati) prodotti dalle

istituzioni deputate alla conservazione del patrimonio bibliografico, archivistico e museale o dai progetti di ricerca basati su tali dati.

Diremo che, se con il Web delle origini l'aspetto computazionale è stato dimenticato, o meglio è stato delegato alle applicazioni locali, con il Web semantico il principio della computabilità torna a emergere e affianca al Web, inteso come sistema di rappresentazione di documenti, un nuovo approccio analitico al dato come esperienza finalizzata alla *knowledge discovery*.[82]

In conclusione, ragionare in termini di semantica (o di Web semantico), nell'orizzonte della costruzione di grafi della conoscenza, significa acquisire consapevolezza di una serie di istanze, che vogliamo sintetizzare in quelli che saranno i prossimi capitoli del volume:

- la potenziale soggettività dell'atto interpretativo di chi ha il compito di descrivere dati culturali a scopo scientifico (cap. 2);
- l'importanza dei contesti per attribuire il giusto significato alle occorrenze e arrivare alla conoscenza (cap. 3);
- la presenza di vari livelli di analisi delle risorse informative, ciascuno dei quali veicola una diversa nozione di semantica (cap. 4);
- la possibilità di valorizzare la semantica attraverso sistemi di visualizzazione in dimensione *user-oriented* (cap. 5);
- la presa di coscienza che la semantica è parte integrante del *workflow* necessario alla realizzazione di progetti nel settore delle DH (cap. 6).

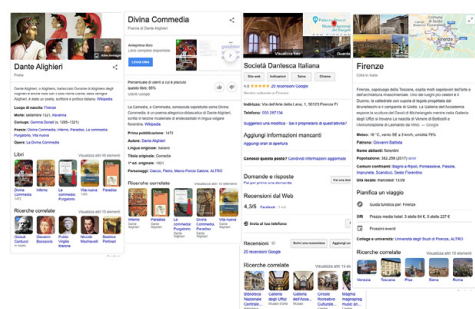

Figura 2 - Una semplice vista lato utente di quanto emerge nel navigare il *knowledge graph* di Google attraverso le *infobox*. Ci si può muovere dal 1) profilo (strutturato) di Dante (persona con ruolo poeta)

collegato ai libri di cui è oggetto, soggetto, autore (edizioni delle opere), alla 2) Commedia (opera) e le risorse correlate (altre opere), fino alla 3) Società dantesca (istituzione), e quindi la 4) città di Firenze (luogo). Quattro entità in potenziale collegamento in un grafo della conoscenza, che aprono, a loro volta, altri collegamenti ad altre risorse di contesto

4 Per la situazione italiana, si può leggere la bibliografia relativa ai contributi scientifici della rivista "Umanistica Digitale", https://umanisticadigitale.unibo.it, accessibile dalle pagine di AIUCD, http://www.aiucd.it/pubblicazioni, e consultabile su Zotero, https://www.zotero.org/groups/1624882/umanistica_digitale/collections/WU6PZRWQ. A livello internazionale un'eccellente progetto è quello di DARIAH (la Digital Research Infrastructure for the Arts and Humanities), dal titolo *Doing Digital Humanities*, che, basandosi sulla tassonomia delle attività delle DH elaborata in seno al progetto TaDiRAH (si veda anche cap. 6), acronimo per Taxonomy of Digital Research Activities in the Humanities, http://tadirah.dariah.eu/vocab, ha collezionato un'importante bibliografia sul tema DH, organizzando le risorse per azioni e scopi, https://www.zotero.org/groups/113737/doing_digital_humanities_-_a_dariah_bibliography/library. Per le origini della disciplina, si veda in particolare Giovanni Adamo, *Bibliografia di Informatica umanistica*, Roma, Bulzoni, 1994.

5 Il riferimento va senz'altro a Tito Orlandi, *Informatica umanistica*, Roma, La Nuova Italia Scientifica, Studi superiori 78, 1990.

6 Il riferimento è innanzitutto a Willard McCarty, *Humanities Computing*, London and New York, Palgrave, 2005. Ma si veda anche, per i principi a fondamento della disciplina, John Unsworth, *What Is Humanities Computing and What Is Not?*, "Jahrbuch für Computerphilologie", 4 (2002), http://hdl.handle.net/2142/157. Per una visione storica e d'insieme si può leggere Susan Hockey, *Electronic texts in the humanities: principles and practice*, Oxford, Oxford University Press, 2000. Ma non dimentichiamo che, anche di recente, c'è chi continua a ragionare in termini di computabilità delle discipline umanistiche come base per le DH. Si veda, ad es., Julianne Nyhan, Andrew Flinn, *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Springer Open, 2018.

7 Un interessante ventaglio di posizioni si può leggere nel capitolo *Selected definitions from the Day of Digital Humanities: 2009-12*, in *Defining digital humanities: a reader*, edited by Melissa Terras, Julianne Nyhan, Edward Vanhoutt, London, Routledge, 2013, p. 279-287. Per avere qualche numero, si vedano le infografiche su DH 2011 dal blog di Melissa Terras, *Quantifying*

Digital Humanities, https://melissaterras.org/2012/01/20/infographic-quantifying-digital-humanities, 2012.

8 *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, John Unsworth, Oxford, Blackwell, 2004, http://www.digitalhumanities.org/companion.

9 *Computing in Humanities Education: A European Perspective*, ACO*HUM, edited by Koenraad De Smedt et al., University of Bergen, HIT-centre, 1999, http://korpus.uib.no/humfak/AcoHum/book.

10 Si possono consultare gli articoli della rivista, fin dalla prima uscita del 1966 su Jstor, https://www.jstor.org/journal/comphuma. Dal 2004 la rivista ha cambiato denominazione ed è diventata "Language Resources and Evaluation" (2005-2016), https://www.jstor.org/journal/langresoeval.

11 Association for Literary and Linguistic Computing (ALLC), dal 2012 ha preso il nome di EADH, la già menzionata Associazione europea.

12 Association for Computer in the Humanities (ACH), https://ach.org., ACH è fra i membri fondatori di ADHO, la già citata Associazione internazionale di DH, nata nel 2005.

13 Dal 2015 la rivista ha cambiato nome, ora è "Digital Scholarship in the Humanities", https://academic.oup.com/dsh, per rimarcare quell'esigenza di superamento di un approccio strettamente legato al dominio della linguistica e della letteratura, e abbracciare una dimensione più ampia della nozione di *humanities* (sulla storia del cambiamento del nome, si veda Edward Vanhoutte, *The Journal is dead, long live The Journal!*, 2015, https://academic.oup.com/dsh/pages/DSH_name_change).

14 *Almanacco letterario Bompiani 1962: le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura*, a cura di Sergio Morando, Milano, Bompiani, 1961.

15 Manfred Thaller, *Defining humanities computing methodology*, in *Computing in Humanities Education: A European Perspective*, cit., ch 2.3.

16 La slide è tratta dal *keynote speech* di Dino Buzzetti, *L'informatica Umanistica fra continuità e trasformazione*, in occasione di AIUCD2019, Ottavo convegno annuale, *Didattica e ricerca al tempo delle Digital Humanities / Teaching and research in Digital Humanities* (Udine, 23-25 gennaio 2019). Il book of abstract si può leggere in *AIUCD2019 - Book of Abstracts*, Quaderni di Umanistica Digitale, 2019, https://doi.org/10.6092/unibo/amsacta/6361.

17 Il fondo Roberto Busa è stato donato nel 2010 all'Università Cattolica di Milano, per cui si veda https://biblioteche.unicatt.it/milano-fondi-fondo-roberto-busa.

18 Roberto Busa, *The annals of humanities computing: The index thomisticus,* "Computers and the Humanities", 14 (1980), 2, p. 83-90. Ora il corpus è online Corpus Thomisticum, http://www.corpusthomisticum.org, 2000-2019. Ma si veda anche il progetto Index Thomisticus Treebank, https://itreebank.marginalia.it, avviato nel 2006 dal centro CIRCSE

dell'Università Cattolica del Sacro Cuore di Milano.
19 Particolarmente significativo allora è l'articolo Marco Passarotti, Geoffrey Rockwell, *The Index Thomisticus as a Big Data Project*, "Umanistica Digitale", 5 (2019), http://doi.org/10.6092/issn.2532-8816/8575.
20 Il termine compare per la prima volta in Franco Moretti, *Conjectures on World Literature*, "New Left Review", 1 (2000).
21 Si veda almeno Teresa Numerico, *Alan Turing e l'intelligenza delle macchine*, Milano, Franco Angeli, 2004.
22 Il riferimento è a Vannevar Bush, *As We May Think*, "The Atlantic Monthly", 176 (1945), 1, p. 101-108.
23 Theodor H. Nelson, *A File Structure for The Complex, The Changing and the Indeterminate*, in *Proceedings of ACM 20th National Conference*, New York, ACM, 1965, p. 84-100.
24 Per una storia dell'automazione dei cataloghi in biblioteca e la nascita del concetto di Online Public Access Catalog, si veda almeno Richard De Gennaro, *Library automation & networking perspectives on three decades*, "Library journal", 108 (1983), 7, p. 629-635.
25 Sul MAchine Readable Cataloguing (MARC) e sulla famiglia dei formati derivati si veda il sito ufficiale della Library of Congress, https://www.loc.gov/marc.
26 Edgar F. Codd, *A relational model of data for large shared data banks*, "Communications of the ACM", 13 (1970), 6, p. 377-387.
27 Harold Borko, *Information science: What is it?*, "American Documentation", 19 (1968), 1, p. 3-5.
28 Si veda, ad es., Andrew Dillon, *What It Means to Be an ISchool*, "Journal of Education for Library and Information Science", 53 (2012), 4, p. 267-273.
29 Il testo completo di quella che oggi è nota come la Classificazione Decimale Dewey è *A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library (Dewey Decimal Classification)*, 1876. Il testo si può scaricare dal sito del progetto Gutenberg, all'indirizzo https://www.gutenberg.org/ebooks/12513.
30 Henry E. Bliss, *A modern classification for libraries, with simple notation, mnemonics, and alternatives*, "Library Journal", 35 (1910), p. 351-358.
31 Paul Otlet, *Traité de documentation. Le livre sur le livre. Théorie et pratique*, Bruxelles, Editiones Mundaneum, 1934.
32 http://www.mundaneum.org.
33 La Classificazione Decimale Universale viene pubblicata per la prima volta con il titolo *Manuel du Repertoire Bibliographique Universel*, Brussels, IIB, 1905-1907; e quindi come UDC *Classification Décimale Universelle*, (FID 151), Brussels, IIB, 1927-1933.
34 Shiyali Ramamrita Ranganathan, *Colon Classification* [First edition], Madras, Madras Library Association, 1933.
35 Una bella panoramica in Michele Santoro, *Ripensare la CDU. Per una*

riflessione sulla storia, il ruolo e le prospettive della Classificazione Decimale Universale, "Biblioteche oggi", 13 (1995), 8, p. 48-57; edizione in "AIB-WEB Contributi" col permesso dell'editore originario https://www.aib.it/aib/contr/santoro1.htm.

36 Jacques Froger, *La critique des textes et son automatisation*, Paris, Dunod, 1968.

37 *La pratique des ordinateurs dans la critique des textes*, a cura di Jean Irigoin, Gian Piero Zarri, Colloque International du CNRS nr. 579, Paris, Editions du CNRS, 1979. La raccolta si apre proprio con le riflessioni di Froger su Quentin: *La méthode de Dom Quentin, la méthode des distances et le problème de la contamination*, p. 13-22.

38 Sul rapporto fra informatica e filologia, nella letteratura contemporanea su questi temi, si veda in particolare Lorenzo Perilli, *Filologia Computazionale*, Roma, Accademia Nazionale dei Lincei, 1995.

39 Si vedano le critiche di Froger al metodo neo-lachmaniano di Henri Quentin, descritto nel suo *Essais de critique textuelle (Ecdotique)*, Parigi, Picard, 1926.

40 "The Thesaurus Linguae Graecae® (TLG®) is a research program at the University of California, Irvine. Founded in 1972 the TLG has collected and digitized most literary texts written in Greek from Homer to the fall of Byzantium in AD 1453. Its goal is to create a comprehensive digital library of Greek literature from antiquity to the present era", http://stephanus.tlg.uci.edu/tlg.php.

41 L'Oxford Text Archive (OTA) viene fondato nel 1976 da Lou Burnard e Susan Hockey sotto l'egida degli Oxford University Computing Services e ancora oggi si presenta come uno dei più sofisticati e corposi *repositories* di testi digitali, accessibili in diversi formati, https://ota.bodleian.ox.ac.uk.

42 "The Packard Humanities Institute (PHI) is a non-profit foundation dedicated to archaeology, music, film preservation, and historical archives". Per l'accesso ai testi latini, https://latin.packhum.org.

43 Si veda in particolare Maurizio Lana, *Biblioteche digitali. Un'introduzione*, Bologna, Bononia University Press, 2013 e ancora Id., *Digital humanities e biblioteche*, "AIB studi", 59 (2019), 1-2, https://doi.org/10.2426/aibstudi-11862.

44 Standard Generalized Markup Language (SGML) è il metalinguaggio di markup dichiarativo, divenuto, dopo una primissima versione come GML, standard dell'ISO nel 1986 (ISO 8879:1986 SGML).

45 La Text Encoding Initiative (TEI) nasce come consorzio per stabilire un modello standardizzato di codifica del testo in formato digitale. Tutte le informazioni sulla pagina ufficiale, http://tei-c.org.

46 La storia della TEI (e delle sue versioni) si legge in *The TEI Archive*, https://tei-c.org/Vault. L'ultima versione delle Guidelines è la *release* P5, rilasciata nel 2007 e costantemente aggiornata. *Guidelines for Electronic Text Encoding and Interchange*, edited by TEI Consortium, http://www.tei-c.org/

P5.

47 Non è questa la sede per una disamina della letteratura sull'argomento ma si vedano, in particolare, le riflessioni intorno al problema delle *overlapping hierarchies* legate al modello OHCO (*Ordered Hierarchy of Content Objects*) in Steven J. DeRose, David G. Durand, Elli Mylonas, Allen H. Renear, *What is Text, Really?*, "Journal of Computing in Higher Education"*,* 1 (1990), 2, p. 3-26. E ancora Allen H. Renear, Elli Mylonas, David G. Durand, *Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies*, prima versione 1993, seconda versione comparsa nella serie *Research in Humanities Computing* 4, edited by Nancy Ide and Susan Hockey, Oxford, Oxford University Press, 1996, p. 263-280. Alcune considerazioni intorno al problema del modello gerarchico ed *embedded* di TEI si possono leggere in Dino Buzzetti, *Digital representation and the text model*, "New Literary History", 33 (2002), p. 61-88; Dino Buzzetti, Jerome McGann, *Critical editing in a digital horizon*, in *Electronic textual editing,* edited by John Unsworth, Katherine O'Brien O'Keeffe, Lou Burnard, New York, MLA of America, 2006, p. 53-73; Domenico Fiormonte, *Scrittura e filologia nell'era digitale*, Torino, Bollati Boringhieri, 2003.

48 Sui primi software per l'inventariazione in archivio si veda il numero monografico della rivista "Archivi & Computer", 13 (2003), 3, *Un'indagine sui programmi di inventariazione archivistica* e in particolare l'introduzione di Stefano Vitali, p. 7-11.

49 Si veda in particolare, circa la storia del Servizio Bibliotecario Nazionale (SBN), il volume dell'Associazione Italiana Biblioteche (AIB), *La cooperazione: il Servizio bibliotecario nazionale*, a cura di Antonella Aquilina D'Amore. Atti del 30 congresso nazionale dell'Associazione italiana biblioteche (Giardini-Naxos, 21-24 novembre 1982), Messina, Università di Messina, Facoltà di lettere e filosofia, Centro studi umanistici, 1986.

50 *International Standard for Archival Description* (ISAD) è standard dell'ICA, ora alla sua seconda edizione, *ISAD(G): general international standard archival description*, adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999, Ottawa 2000. Le linee guida si possono leggere alla pagina ufficiale https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition; trad. it. a cura di Stefano Vitali, con la collaborazione di Maurizio Savoja, in «Rassegna degli Archivi di Stato», 63 (2003), 1, p. 59-190.

51 Encoded Archival Description (EAD), https://www.loc.gov/ead.

52 *Functional Requirements for Bibliographic Records* (FRBR), *Final Report* IFLA, München, K.G. Saur, 1998. La versione corrente è quella con revisioni del 2009, https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf; trad. it., Roma, ICCU, 2000.

53 eXtensible Markup Language (XML), standard del W3C, ultima raccomandazione aggiornata (la quinta) del 2008, https://www.w3.org/TR/xml (versione stabile). Tutto il lavoro del consorzio W3C su XML alla

pagina https://www.w3.org/standards/xml/core.
54 John Unsworth, *Forms of Attention: Digital Humanities Beyond Representation*, "a paper delivered at *The Face of Text: Computer-Assisted Text Analysis in the Humanities*, the third conference of the Canadian Symposium on Text Analysis (CaSTA), McMaster University, November 19-21, 2004", https://johnunsworth.name/FOA. Grassetto nostro.
55 Questa tripartizione prende spunto da Willard McCarty, Harold Short, *Mapping the field*, Report of ALLC meeting held in Pisa, April 2002, http://www.allc.org/node/189, ripreso in Sheila Anderson, Tobias Blanke, Stuart Dunn, *Methodological commons: arts and humanities e-Science fundamentals*, "Philosophical Transactions of The Royal Society. A Mathematical Physical and Engineering Sciences", 368 (2010), p. 3779-3796, https://doi.org/10.1098/rsta.2010.0156.
56 La prima versione del WWW prevedeva anche un editor per ipertesti oltre al sistema di browsing. Si può leggere la storia del Web, assieme al racconto delle origini del progetto, in Tim Berners-Lee, Mark Fischetti, *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor*, San Francisco, Harper San Francisco, 1999.
57 Sul rapporto archivi e Web, si veda, in particolare, Federico Valacchi, *I siti web come strumenti per la ricerca archivistica*, "Archivio Storico Italiano", 160 (2002), 3, p. 589-610.
58 Una panoramica, e una storia del rapporto fra automazione e biblioteche, in Fabio Metitieri, Riccardo Ridi, *Biblioteche in rete. Istruzioni per l'uso*, Roma-Bari, Laterza, 2006.
59 Un fondamentale contributo in materia e George P. Landow, *L'ipertesto: tecnologie digitali e critica letteraria*, Milano, Bruno Mondadori, 1998. Si veda anche Federico Pellizzi, *Letteratura biblioteche ipertesti*, Roma, Carocci, 2005.
60 Anche se alcuni software per le edizioni vengono realizzati con l'avvento del Web. Si pensi in particolare alle versioni di Anastasia di Peter Robinson (http://www.sd-editions.com/anastasia), a Collate per la gestione del processo di collazione (ora CollateX), al pacchetto TUSTEP (http://www.tustep.uni-tuebingen.de/tustep_eng.html). Per una panoramica su tool e applicazioni per l'ecdotica si veda Francesca Tomasi, *L'Ecdotica e le tecnologie. Tra applicazioni, principi e un esperimento*, "Anuario Lope de Vega. Texto, literatura, cultura", 20 (2014), p. 84-98, http://dx.doi.org/10.5565/rev/anuariolopedevega.84.
61 Per una visione d'insieme sulle edizioni digitali a base XML/TEI ad oggi disponibili sul Web si veda, oltre al sito ufficiale della TEI stessa https://tei-c.org/activities/projects, i repertori a cura di Peter Andorfer, Ksenia Zaytseva, Greta Franzini, Catalogue of Digital Editions (CDE), 2016, https://dig-ed-cat.acdh.oeaw.ac.at e la recentissima nuova versione di Patrick Sahle [et al.], Digital Scholarly Editions, v 4.06, 2020, http://www.digitale-edition.de, che rimpiazza lo storico catalogo, v. 3.0 del 2008, a cura dello stesso Sahle. Una più globale riflessione in Federico Meschini,

Oltre il libro. Forme di testualità e digital humanities, Milano, Editrice Bibliografica, 2020.

62 Si vedano, in particolare, fra i volumi più recenti, Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods*, Londra, Routledge, 2015; Paola Italia, *Editing Duemila. Per una filologia dei testi digitali*, Roma, Salerno Editrice, 2020.

63 Si vedano Michele Ansani, *Edizione digitale di fonti diplomatiche: esperienze, modelli testuali*, "RM Reti Medievali", 7 (2006), 2, https://doi.org/10.6092/1593-2214/140; Elena Pierazzo, *A rationale of digital documentary editions*, "Literary and Linguistic Computing", 26 (2011), 4, p. 463-477, https://doi.org/10.1093/llc/fqr033.

64 Per uno stato del dibattito sulle edizioni digitali ad inizio anni duemila, si veda il volume *Electronic Textual Editing*, edited by John Unsworth, Katherine O'Brien O'Keeffe, Lou Burnard, cit.

65 Il primo articolo sul Web semantico è Tim Berners-Lee, James Hendler, Ora Lassila, *The Semantic Web*, "Scientific American", 284 (2001), 5, p. 34-43.

66 Detto anche *Semantic Web Layer Cake*. Si vedano in particolare le attività del consorzio W3C: la W3C Semantic Web Activity, https://www.w3.org/2001/sw, sostituita, a partire dal 2013, dalla W3C Data Activity, *Building the Web of Data*, https://www.w3.org/2013/data.

67 Per il rapporto fra il Semantic Web e la sua realizzazione con i Linked Open Data, si veda *The Semantic Web - Not a piece of cake...* Posted on 2009-07-08 at 14:55 UTC by Benjamin Nowack, e in particolare l'immagine http://bnode.org/media/2009/07/08/semantic_web_technology_stack.png.

68 Terse RDF Triple Language (Turtle) è uno dei possibili formati (.ttl) per esprimere triple RDF, ed è una raccomandazione del W3C, https://www.w3.org/TR/turtle; Resource Description Framework in Attributes (RDFa) è un sistema per incorporare *statements* RDF in HTML, ultima raccomandazione W3C, https://www.w3.org/TR/2015/NOTE-rdfa-primer-20150317; µF o Microformats è un'estensione al markup (attributi @rel, @rev e @class) per aggiungere espressioni semantiche in una pagina HTML (o XHTML), http://microformats.org.

69 Ontology Web Language (OWL), standard del W3C per scrivere ontologie, https://www.w3.org/OWL; Resource Description Framework Schema (RDFS), estensione di RDF con alcune classi e proprietà comuni, https://www.w3.org/TR/rdf-schema; Simple Knowledge Organization System (SKOS), è una famiglia di linguaggi formali creata per rappresentare glossari, classificazioni, tassonomie e qualsiasi tipo di vocabolario strutturato, https://www.w3.org/2004/02/skos.

70 L'ultimissima versione di HTML accessibile sul Web documenta il passaggio dalle canoniche raccomandazioni del W3C al nuovo modello gestito da WHATWG (Web Hypertext Application Technology Working Group), *HTML. Living Standard* — Last Updated 9 November 2021, https://

html.spec.whatwg.org.
71 La prima definizione di Linked Data si può leggere sulle pagine del W3C, e in particolare in Tim Berners-Lee, *Linked Data*, 2006, https://www.w3.org/DesignIssues/LinkedData.html.
72 W3C data activity, *Building the Web of Data*, https://www.w3.org/2013/data.
73 Per una disamina del concetto di Web 2.0 si vedano le pagine del suo ideatore, Tim O' Reilly, *What is Web 2.0*, 2005, http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.
74 Per l'Italia si può vedere il lavoro che sta portando avanti il Ministero con il progetto dati.beniculturali.it, ovvero "la piattaforma in cui il MiBACT [a oggi MiC, n.d.a.] pubblica il proprio patrimonio informativo secondo la logica dei linked open data (LOD). I primi LOD pubblicati rappresentano il frutto di un processo di cooperazione tra gli Istituti centrali e le Direzioni generali del Mibact e collegano tra loro dataset provenienti da fonti diverse: banca dati dei Luoghi della cultura; anagrafiche di Archivi e Biblioteche; banca dati del Catalogo dei beni culturali; altre banche dati documentali e fotografiche", https://dati.beniculturali.it. Il progetto utilizza come ontologia il progetto ArCo "the Italian Cultural Heritage knowledge graph, consisting of a network of seven vocabularies and 169 million triples about 820 thousand cultural entities. It is distributed jointly with a SPARQL endpoint, a software for converting catalogue records to RDF, and a rich suite of documentation material (testing, evaluation, how-to, examples, etc.). ArCo is based on the official General Catalogue of the Italian Ministry of Cultural Heritage and Activities (MiBAC) - and its associated encoding regulations - which collects and validates the catalogue records of (ideally) all Italian Cultural Heritage properties (excluding libraries and archives), contributed by CH administrators from all over Italy". Valentina Anita Carriero [*et al.*], *ArCo: the Italian cultural Heritage Knowledge Graph*, in *International Semantic Web Conference*, Cham, Springer, p. 36-52. In particolare i diversi moduli dell'ontologia del progetto ArCo, assieme ai dati e ad alcuni esempi di query SPARQL, si possono consultare sul sito dedicato, http://wit.istc.cnr.it/arco.
75 Musei, Archivi, Biblioteche (MAB), sito ufficiale http://www.mab-italia.org. Il progetto MAB nasce nel 2011 con l'obiettivo di creare un coordinamento permanente fra AIB (Associazione Italiana Biblioteche), ANAI (Associazione Nazionale Archivistica Italiana) e ICOM Italia (International Council of Museum - Comitato Nazionale Italiano) al fine di "esplorare le prospettive di convergenza tra i mestieri e gli istituti in cui operano i professionisti degli archivi, delle biblioteche, dei musei".
76 Libraries, Archives, Museums (LAM), più di recente ribattezzato in Galleries, Libraries, Archives, Museums (GLAM), termine che non nasce come vero coordinamento fra istituzioni, ma per riferirsi in generale agli istituti culturali. Uno dei primi report OCLC sulle azioni di collaborazione

negli istituti americani si può leggere in Diane Zorich, Günter Waibel, Ricky Erway, *Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums*, 2008, https://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf.

77 Un significativo passo in avanti sarà rappresentato dai lavori del neonato Istituto centrale per la digitalizzazione del patrimonio culturale – Digital Library che ha "l'obiettivo di coordinare e promuovere i programmi di digitalizzazione del patrimonio culturale del Ministero della Cultura […] Inoltre, la Digital Library svolge le funzioni di indirizzo e controllo sull'Istituto centrale per gli archivi, l'Istituto centrale per i beni sonori e audiovisivi, l'Istituto centrale per il catalogo e la documentazione e l'Istituto centrale per il catalogo unico delle biblioteche italiane", https://digitallibrary.cultura.gov.it.

78 Si veda in particolare una recente monografia che raccoglie gli atti di un convegno su quelle edizioni digitali che utilizzano il modello di dati a grafo e le tecnologie del Web semantico per la valorizzazione della conoscenza: *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, Georg Vogeler, Norderstedt, Books on Demand, 2021. Una riflessione sul rapporto XML/TEI vs RDF in Georg Vogeler, *Digital Edition of Archival Material - Machine Access to the Content: On the Role of Semantic Web Technologies in Digital Scholarly Editions*, in *Digitizing Medieval Sources? L'édition en ligne de documents d'archives médiévaux*, edited by Christelle Balouzat-Loubet, Turnhout, Brepols, 2020, p. 37-56, https://doi.org/10.1484/M.ARTEM-EB.5.117327.

79 Amit Singhal, *Introducing the Knowledge Graph: Things, Not Strings*, Google Official Blog, May 16, 2012, https://blog.google/products/search/introducing-knowledge-graph-things-not. Ma si veda anche Lisa Ehrlinger, Wolfram Wöß, *Towards a definition of knowledge graphs*, "SEMANTICS" (Posters & Demos Track), 48 (2016), p. 1-4.

80 DBpedia, https://www.dbpedia.org.

81 Wikidata, https://www.wikidata.org/wiki. Sull'importanza crescente di questo dataset, si veda Denny Vrandečić, Markus Krötzsch, *Wikidata: A Free Collaborative Knowledgebase*, "Communications of the ACM", 57 (2014), 10, p. 78-85, https://doi.org/10.1145/2629489.

82 Si veda, ad es., Eero Hyvönen, *Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery*, "Semantic Web", 11 (2020), 1, p. 187-193, https://doi.org/10.3233/SW-190386.

2. DESCRIZIONE O INTERPRETAZIONE? IL RUOLO DELLE DH NELLA SFIDA ERMENEUTICA

La documentazione è l'azione mentale
che analizza e interpreta il contenuto
concettuale del documento al fine di individuarne le unità informative e diffonderle con apposite procedure.[83]

2.1 Digital Humanities e Library and Information Science

Quando si parla di scienza dell'informazione (*information science*) si allude a un tema dai confini spesso sfumati o anche fraintesi dalla comunità. Il termine con cui oggi, in taluni contesti, ci si riferisce alle scienze dell'informazione come alla sola componente informatica, rischia di cancellare la storia di una disciplina che si concentra invece sugli strumenti, i metodi e le tecniche per una gestione efficace ed efficiente dell'intero ciclo vitale dell'informazione, cioè del dato arricchito di contesto. Detta con Borko:

> Information science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the **origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information**. […]. It is an

> interdisciplinary science derived from and related to such fields as **mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, management, and other similar fields**. It has both a pure science component, which inquires into the subject without regard to its application, and an applied science component, which develops services and products.[84]

Da questa definizione deduciamo due concetti molto importanti: il primo è riconoscere l'*information science* come quella disciplina che lavora sul trattamento dell'intero ciclo vitale dell'informazione e che quindi detiene il primato della trasversalità, che è ovvero capace di governare sistemi di gestione, manipolazione e valorizzazione dell'informazione; l'altro è individuare una varietà di settori disciplinari diversi che contribuiscono a elaborare la nozione di *information science*, apportando ciascuno una fetta di conoscenza necessaria a stabilire e determinare il senso di questa disciplina. Un sapere nuovo dunque, che si forma dell'esperienza di campi di ricerca diversi, e che nasce dall'interdisciplinarità – che non è sommatoria di saperi, ma trasversalità –, così come accaduto per le DH.

Gestire l'informazione, o anche organizzarla, è l'attività che da sempre qualifica l'operato di archivi, biblioteche e musei.[85] Il passaggio alla LIS,[86] anche se non uniformemente condiviso dalla comunità scientifica,[87] è dunque naturale. Esso arricchisce ulteriormente la riflessione, aggiungendo le pratiche delle biblioteche e della documentazione alla gestione del ciclo vitale dell'informazione.

La cosa che qui più ci interessa è l'inevitabile serie di punti di contatto, dialogo e interconnessione della LIS con le DH.[88]

Innanzitutto, ci sono alcuni temi, cari alla LIS,[89] che potremmo sintetizzare in una serie di parole chiave (le riportiamo in ordine alfabetico):

- *bibliometrics*
- *digital libraries*
- *digital preservation*
- *information architecture*
- *interaction design*
- *scholarly communication*
- *user experience*

Questi temi possono essere considerati i primi punti di contatto,

o anche i primi ambiti di un comune orientamento della ricerca scientifica, per quanto riguarda: gli aspetti di progettazione, creazione e modellazione di una raccolta di oggetti digitali (il tema della *digital library*[90]), quelli di implementazione, sia a livello teorico (la *scholarly communication*) che a livello tecnico-tecnologico (l'uso di tecniche matematico-statistiche per la bibliometria), il tema della preservazione dell'informazione digitale, fino al rapporto con l'utente finale (*user experience* ma anche *information architecture*) e alla messa a disposizione di strumenti di accesso ai contenuti di una collezione (*interaction design*).

Dai temi, allarghiamoci agli ambiti, per estendere il nostro ragionamento. Riflessioni sul rapporto DH e LIS non mancano infatti in letteratura, e contribuiscono ad estendere l'orizzonte del ragionamento. Così scrive Chris Alen Sula:

> A search for "digital humanities" within library and information science literature reveals a steady increase in publications since 2005 in the Library, Information Science & Technology Abstracts (LISTA) database, which indexes over 700 journals as well as books, research reports, and proceedings […]. It is remarkable that publications on digital humanities have nearly doubled in 2012, with more still being indexed at the time of this publication.[91]

E se nell'anno 2012 le pubblicazioni del settore LIS che includono la parola "digital humanities" sono raddoppiate rispetto agli anni precedenti, dal 2012 a oggi la situazione è cresciuta in modo esponenziale.[92]

Ed è indubbio che ci siano teorie, metodi e tecniche che attraversano trasversalmente saperi che nascono interdisciplinari dalla loro stessa formazione. Riprendendo ancora Sula, ci sono cinque macro ambiti, o anche *cluster* o *topics*, di interconnessione profonda tra LIS e DH (Fig. 3):

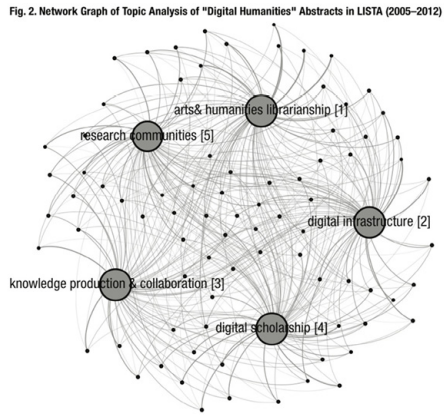
Fig. 2. Network Graph of Topic Analysis of "Digital Humanities" Abstracts in LISTA (2005–2012)
arts& humanities librarianship [1]
research communities [5]
digital infrastructure [2]
knowledge production & collaboration [3]
digital scholarship [4]

Figura 3 - Interconnessioni fra ambiti di ricerca comuni a DH e LIS.
Tratto da http://chrisalensula.org/digital-humanities-and-libraries-a-conceptual-model

- *arts & humanities librarianship*
- *digital infrastructure*
- *knowledge production & collaboration*
- *digital scholarship*
- *research communities*[93]

Questi temi sono assolutamente traversali rispetto ai due domini e confermano linee di ricerca condivisa: l'interesse per il settore umanistico, la riflessione sui temi infrastrutturali, il concetto di conoscenza e le modalità della sua gestione, la cultura (del) digitale come metodologia, e l'esigenza di costruire delle comunità di ricerca scientifica.

Questa riflessione si aggiunge quindi alla classificazione per parole chiave sopra riportata, restituendo un approccio per ambiti, che si può allargare ulteriormente in un approccio per azioni.

Gli ambiti appena descritti corrispondono infatti ad altrettante azioni, comuni tanto a LIS quanto a DH:

- *managing data*
- *'embedded' librarianship*
- *digitization and curation*
- *digital preservation*
- *discovery*
- *dissemination*[94]

Di nuovo, cioè, l'importanza del ciclo vitale: produzione, manipolazione, preservazione, disseminazione e scoperta.[95]

Temi, ambiti e azioni appena descritti determinano dunque quello spazio del dialogo disciplinare che avvicina le DH alla LIS. Ma se dovessimo individuare, rispetto a questo approccio tridimensionale, la caratteristica che meglio qualifica le DH in un contesto fatto di trasversalità dovremmo spostarci sul piano dell'ermeneutica, ovvero anche del ruolo di una disciplina che individua nel processo di interpretazione il suo primato. Ciò che le DH hanno da dare è il ragionamento critico sui dati culturali, esprimendolo nei termini di quesiti di ricerca, finalizzati a rispondere a specifiche esigenze scientifiche; lavorare cioè sui dati per creare nuovi dati, che significa trasformare risorse culturali in strumenti per veicolare una conoscenza nuova.

Così come già altrove proposto:

> Gli àmbiti di ricerca che le DH naturalmente accolgono vanno oltre il tradizionale connubio fra discipline umanistiche e informatiche, coinvolgendo tutta la componente giuridica, sociologica, economica e statistica, e superando la formale complementarietà, in favore di una necessaria ibridazione. Ciò che le DH hanno da offrire ad un simile dibattito è un patrimonio di pratiche e ragionamenti che potrebbero trasformare la progettualità nata nell'alveo di una disciplina tradizionale in nuove domande di ricerca, arrivando potenzialmente a raggiungere risultati non preventivati e non altrimenti determinabili. Se volessimo riassumere una visione del ruolo delle DH, sicuramente la prospettiva di svelare l'inaspettato e far emergere il non conosciuto rappresenterebbe l'obiettivo forte di questo àmbito di ricerca.[96]

Descrivere oggetti culturali e standardizzare le procedure, attraverso vocabolari condivisi dalle comunità, sono il primo passo verso la rappresentazione formale dei dati. Senz'altro il primato della descrizione unisce LIS a DH proprio perché la descrizione è l'azione che sta a fondamento di ogni processo di gestione dei dati. Aggiungere il piano dell'interpretazione significa introdurre il tema della soggettività, della molteplicità, della stratificazione e del potenziale *disagreement* e quindi della possibilità di ottenere da quei dati nuove risposte a nuovi quesiti.[97] Ecco che la descrizione delle risorse culturali dovrà essere in grado di raccontare la storia dei dati, di guidare il fruitore in un percorso di apprendimento dell'evoluzione del bene nel suo rapporto con chi quel bene ha descritto, conservato, divulgato, ma anche usato e studiato. La descrizione deve cioè avere il compito di essere una narrazione storica, capace di documentare dati e contesti. Ne consegue che non

dovranno essere i dati ad adeguarsi necessariamente alle specifiche stabilite dagli standard, ma gli standard dovranno essere ripensati rispetto al ruolo che ricoprono, nell'ottica di favorire la ricerca scientifica, aprendo la descrizione degli oggetti culturali al piano dell'interpretazione, come esito dello studio su quegli stessi dati. La ricerca scientifica sui dati del patrimonio culturale si traduce naturalmente nella scoperta di nuovo sapere, che può valorizzare quello stesso patrimonio culturale aggiungendo altri elementi utili per l'avanzamento delle conoscenze sui beni e sulla loro storia.

Mi piace riportare, e condividere con il lettore, un recente commento di Dino Buzzetti a un mio articolo scientifico:

> Farei due osservazioni: (a) la difficoltà nello stabilire e nel classificare gli standard dipende, come si è detto, dal fatto che la semantica è indeterminata in linea di principio e che può essere solo relativa, ossia stabilita in relazione a un determinato contesto; (b) la relazione tra 'data' e 'metadata', ovvero dati e metadati, è il punto centrale. Mi sono già espresso sulla formalizzazione dell'interpretazione che è propria dell'autore del progetto. A mio modo di vedere il rapporto tra un'interpretazione propria e la descrizione standard dovrebbe essere rovesciato. Mi spiego: non si tratta di adattare la propria interpretazione agli standard, ma di fondare e ricavare gli standard dall'analisi dei dati, ossia dalla produzione dei LOD come risultato delle pratiche analitiche di ricerca. A quel punto sono gli standard che vanno usati in modo tale poter essere adattati ai LOD ricavati dai dati. È banale: è la scarpa (i metadati) che deve adattarsi al piede (i dati) e non il piede alla scarpa. Di qui anche il rapporto tra ricerca (ossia elaborazione e produzione di dati) e pratica di metadatazione (ossia di descrizione dei dati prodotti). Non sono contrario ai linguaggi controllati e agli standard (ci mancherebbe!) che sono necessari per rendere interoperabili i dati di ricerca, ma i linguaggi controllati e gli standard dovrebbero essere fondati sui risultati della ricerca e non si dovrebbero modificare i dati di ricerca per poterli adattare costrittivamente alle forme previste dagli standard.

Ed è con queste eloquentissime parole che vogliamo introdurre il presente capitolo, con l'obiettivo di rispondere a queste osservazioni, in un percorso che dal problema della descrizione delle risorse culturali attraverso sistemi standardizzati ci consenta di arrivare al tema dell'interpretazione di quei dati. Perché il cuore delle DH è risolvere uno speciale bisogno di ricerca, ovvero rispondere a precisi obiettivi conoscitivi attraverso la modellazione del dominio di analisi. Se in campo MAB/(G)LAM si producono dati e metadati, in ambito DH si usano quei dati e metadati per generare nuova conoscenza.

2.2 Descrivere e standardizzare

Descrivere risorse culturali è l'attività che da sempre contraddistingue l'operato delle istituzioni deputate alla conservazione del patrimonio, di quel patrimonio che rappresenta la memoria condivisa e collettiva.[98]

Riflettendo in particolare sulla situazione italiana, non c'è dubbio che ICCU,[99] ICAR[100] e ICCD[101] siano i tre pilastri su cui storicamente si basa l'esigenza di definire e condividere le buone pratiche per la descrizione delle risorse – eterogenee, sfaccettate e multiformi –, che popolano l'orizzonte culturale, codificato dalla storia della tradizione materiale. Confrontandosi con il panorama internazionale dell'IFLA,[102] dell'ICA[103] e dell'ICOM,[104] l'Italia ha elaborato una propria serie di standard, metodologici prima e strutturali poi, fino a costruire modelli concettuali ad hoc. La prospettiva è anche quella dell'integrazione – come testimoniato dall'attività del MAB – e il fine è di affrontare il tema del dialogo fra modelli descrittivi spesso davvero divergenti a fronte di artefatti oggettivamene eterogenei (dalla fotografia all'opera d'arte, dal manoscritto al libro a stampa, dal sigillo alla mappa cartografica, dal periodico al libro antico, solo per fare qualche esempio).

Da MARC a ISBD,[105] dalle AACR – AACR2[106] al più globale standard rappresentato dal nuovo modello, anche teorico, RDA[107] per il mondo digitale, e ancora dal modello concettuale veicolato da FRBR e dalla sua famiglia FRAD[108] e FRSAD,[109] più di recente ripensati come un modello unico IFLA LRM,[110] fino all'ontologia americana BIBFRAME,[111] le biblioteche hanno riservato larga parte della propria riflessione teorica a standardizzare il processo di costruzione del catalogo: dal rilascio di linee guida teoriche fino all'elaborazione di modelli ontologici.[112] In un'ottica che da sempre condivide con gli archivi l'esigenza di separare la descrizione dell'oggetto – tipicamente al livello dell'item –, o anche il record catalografico (ora i dati), dalla descrizione della responsabilità intellettuale (persona, ente o famiglia) a cui ascrivere una funzione di paternità a un qualche livello, sull'oggetto stesso. Alla descrizione si aggiungere il tema della classificazione assieme a quello della soggettazione, azioni tipicamente demandate agli standard per la gestione dei valori associati ad alcuni specifici elementi della descrizione. Andiamo allora dai già citati sistemi di classificazione decimale DDC[113] e UDC,[114] assieme alla CC,[115] fino

ai sistemi di soggettazione, come LCSH[116] o, in Italia, il Nuovo soggettario della BNCF.[117]

Un caso a parte rappresenta la descrizione del manoscritto, che ha saputo trovare in TEI un adeguato vocabolario per la modellazione.[118] E, in tema di manoscritti, non possiamo dimenticare il più recente esperimento, rappresentato dal IIIF,[119] che sta diventando un *framework* adottato da sempre più numerose istituzioni di conservazione che vogliano garantire interoperabilità delle immagini delle fonti primarie acquisite digitalmente. In fondo IIIF suggerisce, assieme alle API che il progetto mette a disposizione, un set di metadati strutturali attraverso cui produrre quel *manifest* (ovvero un file in JSON-LD[120]) a garanzia di interscambio.

E gli archivi non sono stati da meno: il già citato standard metodologico ISAD assieme a ISAAR(CPF)[121] e i conseguenti standard strutturali EAD e EAC(CPF),[122] e ancora ISDF[123] e ISDIAH,[124] e il più recente RiC-CM[125] con l'ontologia RiC-O,[126] ma anche le italiane OAD[127] e la EAC(CPF) Ontology,[128] configurano e delineano l'esigenza di condividere principi, modelli e schemi per la rappresentazione dei complessi documentari e dei soggetti produttori, nel solco il principio della descrizione separata.

La situazione dei musei, o anche dei più generali beni culturali, trova nell'operato internazionale dell'ICOM, e dell'italiano ICCD, una proposta che deve fare i conti con una molteplicità di supporti e beni (dai beni storico-artistici a quelli archeologici, dai beni demoetnoantropologici a quelli naturalistici, dai beni architettonici e paesaggistici a quelli fotografici, numismatici, musicali fino ai beni scientifici e tecnologici);[129] questo significa molteplicità di tipologie documentarie, ciascuna delle quali determina un approccio diverso al bene culturale. Certamente però, con l'elaborazione dello standard CIDOC CRM,[130] i musei hanno provato, e sono in parte riusciti, a proporre un modello a cui tutti gli istituti culturali stanno progressivamente conformandosi.[131]

E poi ci sono, giusto a titolo di esempio, DC[132] e DC Terms,[133] l'ontologia FOAF,[134] Schema.org,[135] e tutti quegli standard trasversali che stanno contribuendo ad aumentare esponenzialmente quella serie di modelli di metadatazione cui il nuovo Web sta chiedendo di adeguarsi.[136] Dagli standard, nativamente pensati per descrivere risorse culturali analogiche, le istituzioni culturali si sono dovute confrontare con la descrizione di

oggetti digitali, risultato di un processo di conversione di quelle fonti primarie da risorse tangibili da catalogare e inventariare, a risorse da metadatare.

Se volessimo tentare di classificare o anche di categorizzare gli standard che nel tempo hanno contraddistinto l'approccio delle istituzioni alla descrizione delle risorse culturali, incontreremmo una varietà di posizioni.

Tradizionalmente, quando si parla di metadati, si adotta un approccio tripartito, trasversale rispetto ai domini: metadati descrittivi, tipicamente compito del modello DC; metadati tecnici e amministrativo-gestionali, ruolo che in Italia è ricoperto dallo standard MAG[137] dell'ICCU; metadati strutturali, tradizionalmente rappresentati attraverso il sistema METS,[138] ideato e manutenuto dalla Library of Congress. Ma esistono anche altre posizioni, che aggiungono alla tripartizione le classi dei metadati per la preservazione (in particolare lo standard PREMIS[139]), metadati per la gestione dei diritti e gli stessi linguaggi di markup.[140]

Non mancano posizioni che assumono questa tripartizione (descrittivo, tecnico e strutturale) in un più ampio sistema classificatorio che trova nei concetti di dominio (*domain*), funzione (*function*), scopo (*purpose*) e comunità (*community*) la modalità per stabilire le differenti modalità di approccio alla descrizione e quindi per affrontare l'organizzazione degli standard esistenti.[141]

Altro possibile sistema è quello che lega la classificazione alla funzione. Ecco allora che abbiamo standard di contenuto (*content*) che sono quelli che valorizzano l'approccio in termini di *guidelines* o *best practices* svincolate dalla modalità della serializzazione, di struttura (*structure*), che legano la funzione dello standard alla sua espressività computazionale (DTDs o Schemi), di interscambio (*exchange*), e di codifica o formato (*encoding*) o anche di valore (*value*), vale a dire le forme controllate associabili a certi elementi (Tabella 1).

Più recentemente, ovvero con l'affermazione del paradigma dei LOD, la classificazione degli standard preferisce focalizzarsi sul ruolo formale del metadato, proponendo tre livelli di gestione dei dati: 1) i set di descrittori – ovvero gli Schemi (*element sets*); 2) le ontologie o più in generale i vocabolari RDF – concretizzazione della modellazione concettuale; 3) i valori associati agli elementi (*value vocabularies*) – ovvero i termini acquisiti dai vocabolari controllati.[142] Ecco che: 1) XML non è più l'unico formato di

serializzazione, ma la capacità descrittiva della tripla soggetto-predicato-oggetto, così come delineata dall'RDF, trova altre sintassi per esprimersi (da Turtle a N-Triples, da JSON-LD a RDFa, da Trig a N3, oltre a RDF/XML); 2) gli Schemi sono progressivamente sostituiti dalle ontologie, cioè concettualizzazioni più espressive, che invece di ragionare per elementi, attributi e valori ragionano per classi, predicati e istanze; 3) i valori associati ai metadati non sono più solo elenchi tassonomici di termini di un dizionario controllato, ma sono a loro volta ripensati (e pubblicati) in ottica LOD (basti pensare al database lessicale Wordnet,[143] o a VIAF,[144] all'attività della Library of Congress con il suo Linked Data Service,[145] a Getty con i suoi strumenti,[146] o ancora a progetti come il database di luoghi geografici GeoNames[147] o il *gazetteer* di luoghi antichi Pleiades[148]).

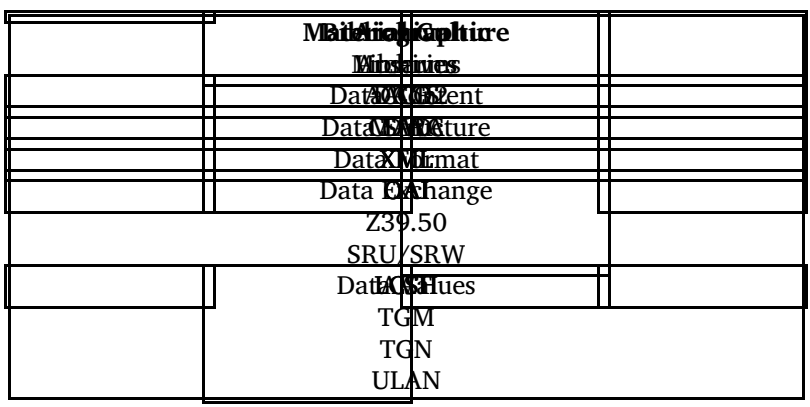

Tabella 1 - La classificazione tradizionale degli standard prima dell'avvento dei LOD. Una personalizzazione dello schema. Tratto da Mary W. Elings, Gunter Waibel, *Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives and Museums*, "First Monday", 12 (2007), 3, http://firstmonday.org/issues/issue12_3/elings/index.html

Progetti come il SAN LOD,[149] CulturaItalia[150] del MiC, Internet Culturale[151] di ICCU assieme al neonato progetto Alphabetica,[152] SigecWeb, ovvero il Catalogo generale dei Beni Culturali[153] con gli open data di ICCD, e i già citati dati aperti del MiC,[154] rappresentano la volontà di raccogliere le descrizioni degli oggetti conservati presso le istituzioni culturali e fornire un servizio di accesso alla descrizione di quegli oggetti, in una prospettiva che sempre di più vede nei LOD lo strumento finale di disseminazione

dei risultati della manipolazione.[155]

Date queste premesse, e al di là dell'aspetto classificatorio che si riflette nelle tipologie di standard e metadati, un'importante riflessione, prima ancora che sui LOD, va fatta sul concetto di oggetto della descrizione, perché è su quello che va condotta ogni azione finalizzata ad aumentare la capacità delle risorse di diventare strumento di conoscenza, soprattutto quando questi oggetti vengano usati a scopi di ricerca scientifica.

Fino a quando si ragioni in un contesto analogico, gli standard di contenuto e di struttura hanno la capacità di determinare, con la scheda catalografica, o più in generale descrittiva, ma anche con gli strumenti di corredo archivistici, la rappresentazione più adeguata di una fonte informativa, attraverso la definizione dei parametri della descrizione e gli elementi per la formalizzazione. Scheda catalografica e strumenti di corredo sono quindi il tramite per l'accesso alla conoscenza veicolata dalle fonti materiali.

Quando ci si voglia invece spostare in ambiente digitale e si acquisisca digitalmente una fonte analogica, è evidente che gli elementi della descrizione si riferiscono a oggetti che naturalmente appartengono a livelli diversi: da un lato la fonte analogica, dall'alto le numerose possibili riproduzioni digitali di quella fonte (in un rapporto potenziale 1:n). E l'appiattimento a cui le categorie costringono (descrivere la fonte analogica non è uguale a descrivere una sua riproduzione digitale) rischia di far perdere importanti informazioni. Ecco che nozioni condivise e consolidate, come quelle di autore, data, luogo, formato o tipo, solo per dirne alcune, vanno ripensate in funzione del livello cui l'elemento descrittivo si riferisce.

Pensiamo alle tradizionali categorie del DC, classico esempio di un già menzionato approccio classificatorio cosiddetto a faccette, le cui limitazioni ha cercato di risolvere DC Terms attraverso l'impiego dei qualificatori e l'estensione delle tipologie di possibili relazioni.

Se si osserva un oggetto analogico, per esempio un libro, e lo si voglia convertire in un formato digitale, e si voglia stabilire il concetto di "data", è naturale domandarsi a quale data ci si riferirà. Sappiamo che il DC mira a descrivere, usando la terminologia FRBR, una speciale "manifestazione" di una "espressione" intesa come particolare contributo intellettuale al processo della creazione. Ma è sufficiente limitare la capacità espressiva di un oggetto culturale alla sua manifestazione materiale? E fin qua si

potrebbe immaginare di usare standard diversi: ad es., al DC il compito di dare informazioni sulla manifestazione a livello descrittivo e magari attribuire a MAG quello di descrivere gli aspetti amministrativo-gestionali. Per poi farli dialogare (ad es. attraverso METS).

Ma andiamo più a fondo e cerchiamo di capire se questo ci basta. Con i metadati si sta descrivendo un oggetto, analogico o digitale che sia, o si sta creando la scheda di quell'oggetto? Se la risposta è la scheda, allora sarà necessario documentare la data di creazione della scheda, o anche la data della sua revisione.

Diverso quindi descrivere l'oggetto culturale come risorsa capace di veicolare un contenuto, potenzialmente plurimo e sfaccettato, descriverne una speciale manifestazione o realizzare la scheda, o lo strumento di corredo più in generale, che lo identifica (a cui aggiungiamo complessità pensando a quelle schede che sono esse stesse fonti storiche). Una "cosa" (*thing*), secondo la terminologia del Semantic Web, è l'oggetto scheda, altra "cosa" è la risorsa analogica, cioè la fonte materiale, e ancora diversa "cosa" è una delle sue possibili riproduzioni digitali (senza peraltro addentrarsi qui sugli ulteriori livelli della stratificazione del processo descrittivo). Abbiamo cioè a che fare con entità differenti, ciascuna delle quali merita di essere documentata.

È evidente che uno standard come ad es. ISBD chiarisce a che "data" ci si vuole riferire quando si realizzi la scheda catalografica, ma in ambiente digitale ci sono varie "date" che andranno tenute in considerazione nel documentare il ciclo di vita dell'oggetto. E per garantire la massima espressività di quella descrizione. Pena la perdita di informazione essenziale a documentare l'azione descrittiva in tutto il suo spessore semantico, necessario a fare dell'oggetto digitale un'entità dotata di un autentico valore culturale.

La considerazione è quindi che la descrizione degli oggetti culturali andrebbe ripensata nella prospettiva di documentare la stratificazione delle fonti (ovvero la loro capacità espressiva), separando contenuto da contenitore, analogico da digitale, scheda descrittiva da artefatto. Con tutte le implicazioni che questo modello determina.

Ma ci sono altri motivi per cui spesso la metadatazione può creare ambiguità agli occhi di chi fruisce di quei dati, ed è per questo un'azione che richiede un approccio critico all'osservazione

del bene da descrivere. Ci possono essere descrizioni obsolete, non aggiornate alla luce di nuove scoperte; possiamo avere dati descrittivi con grado di certezza variabile (per esempio un'attribuzione di paternità, la data di creazione di un contenuto, il riconoscimento di un grafema in un manoscritto); possono essere state fatte delle forzature di una descrizione in un campo non adeguato ad ospitare una certa informazione; possono essere state inserite preziose informazioni in forma non strutturata, rendendole potenzialmente non raggiungibili; ma soprattutto ci possono essere posizioni divergenti sullo stesso oggetto di analisi, ciascuna delle quali meriterebbe di essere documentata. I metadati andranno cioè essi stessi dotati di contesti necessari a garantire la qualità delle asserzioni e quindi assicurare la loro affidabilità. E il primo di questi contesti è la documentazione dell'interprete, inteso come lo studioso che osserva quei metadati per arricchire la sua esperienza conoscitiva su quei beni (ovvero chi fruisce di quelle descrizioni, ma che può anche diventare attivo produttore di nuovi contenuti in ottica di arricchimento).

2.3 Interpretare

Pensiamo a un oggetto complesso, fra i più sfuggenti, come la cartolina, o anche la lettera, che potenzialmente già di per loro sono oggetti complicati dal punto di vista descrittivo. A cosa allude il concetto di data? La data di creazione, si dirà. Ma non saranno altrettanto importanti, dal punto di vista della necessità di documentare il contenuto della fonte, anche la data di ricezione? O la data dell'affrancatura? O la data del timbro postale? O la data di realizzazione del supporto materiale? E siamo solo nella dimensione analogica. Stesse riflessioni condotte, a titolo di esempio, sulla nozione di tempo (la data), potrebbero essere condotte sulla dimensione spaziale (i luoghi), o ancora sull'attribuzione di paternità delle azioni (gli agenti) – e sul ruolo svolto da ogni entità nel processo che governa il ciclo vitale dell'oggetto culturale –, ma anche sulle nozioni di formato o tipo.

Oggetto analogico concettuale, contenuto(i) di quell'oggetto, oggetto fisico, tanto analogico quanto digitale, scheda descrittiva di ciascuno di questi livelli sono le lenti di esplorazione di quelle risorse che contribuiscono a formare l'identità molteplice del nostro

patrimonio culturale.

Alla necessità di riconoscere l'esigenza di una descrizione a livelli, come FRBR insegna (per cui si veda cap. 4), si aggiunge la necessità di poter disporre di modelli formali che, su base gerarchica/tassonomia prima e relazionale/reticolare poi, diano consistenza alla stratificazione dei livelli. Le ontologie in questo senso rappresentano la necessaria concettualizzazione, che deriva dall'analisi degli oggetti di un dominio – o di una raccolta, collezione, ma anche di un fondo o complesso documentario – e delle loro sfaccettature, plurime e trasversali. Sono le ontologie cioè che permettono di dare valore alle entità e anche di fornire spessore alle relazioni, semantizzando le gerarchie e attribuendo i descrittori al corretto livello di osservazione di un oggetto come componente di una raccolta, oltre ad attribuire la giusta classe o proprietà all'informazione da rappresentare.

Le DH hanno il ruolo di ripensare agli Schemi in un'ottica ontologica, ma aggiungendo la lente dell'interprete. Di tradurre elenchi piatti di elementi della descrizione (medatati a semantica implicita) in classi relazionate attraverso proprietà semanticamente esplicite, capaci di tradurre i problemi della ricerca umanistica attraverso la modellazione. Ma anche di concettualizzare l'interpretazione dell'esperto di dominio, che può tradurre gli elementi della descrizione ritenuti pertinenti, in un sapere fruibile tanto dalla macchina quanto dall'utente finale. Perché è vero che l'ontologia di FRBR, ovvero FRBRoo,[156] ci permette di lavorare nella dimensione del modello concettuale, ma il processo di traduzione della semantica implicita, come espressa da etichette descrittive tradizionali degli standard, in semantica esplicita, ovvero le etichette ripensate come classi e predicati, deve essere documentato. Soprattutto quando ragioniamo nell'ottica del Web semantico e ci poniamo il problema di utilizzare un data model (le componenti scelte di una o più ontologie) per creare un dataset (l'insieme degli asserti espressi come *statements* RDF).

Aggiungeremo che l'approccio delle istituzioni culturali, derivato dalla tradizione degli standard, è fondato sul principio della neutralità del punto di vista. Ovvero non è ammessa contraddizione, ma vengono documentate le caratteristiche oggettive delle fonti, certamente attraverso etichette ricche e appropriate, lasciando però poco spazio alla complessità intesa come molteplicità,[157] che sia un contesto analogico o digitale. E su

questo aspetto andrà dedicata un'opportuna riflessione perché ben sappiamo che ogni affermazione su aspetti non fattuali degli oggetti culturali (ancora di più se si lavora sul testo pieno dell'oggetto culturale) è esito di un'osservazione, è risultato di un ragionamento, è l'esito della consultazione delle fonti attraverso cui attribuire caratteristiche all'oggetto dell'analisi. Ed è attraverso un approccio finalizzato a rendere esplicita l'interpretazione che è possibile risolvere alcuni dei problemi legati a una descrizione tradizionale, ovvero neutrale, basata su metadati riferiti ai beni culturali.

Ma andiamo per gradi e pensiamo ancora al problema della descrizione degli oggetti culturali, in un contesto che voglia arrivare alla trasformazione da un sistema tradizionale relazionale a un dataset di triple nell'ottica del Semantic Web. E riflettiamo sul fatto che, nonostante gli sforzi che la comunità MAB/(G)LAM sta facendo per uniformare e riconciliare i modelli descrittivi:[158]

- ogni oggetto, ovvero anche ogni risorsa culturale sulla base della tipologia a cui appartiene, ha i suoi standard di riferimento;
- ogni istituzione usa i suoi standard, e quindi stabilisce il set di descrittori che meglio qualificano la sua attività;
- non è detto che un'istituzione non preferisca piegare le proprie modalità consolidate di descrizione per descrivere oggetti che andrebbero meglio descritti con altri standard.

Come risolvere il dialogo allora fra le istituzioni senza svilire la capacità espressiva dei singoli standard? Domanda a cui i grandi aggregatori, come Europeana, con il suo EDM,[159] hanno tentato di rispondere con modelli di allineamento e di *mapping* fra gli standard rappresentativi di tipologie documentarie diverse e uno schema o modello capace di conciliare punti di osservazione differenti, perché basati su altrettanto diversi oggetti culturali nel dominio MAB/(G)LAM.[160] Se certamente la sfida dell'interoperabilità impone un approccio orientato all'allineamento semantico tra modelli concettuali,[161] sarà lecito chiedersi se sia solo questo il problema. O se non sia invece necessario documentare le attività descrittive arricchendo la rappresentazione delle risorse con gli strumenti in grado di validare la qualità dell'informazione stessa.

Da un lato questo significa attribuzione di paternità. Ogni azione descrittiva va infatti legata all'interprete che quella descrizione ha

fatto. A un primo livello ci riferiamo alla necessaria predisposizione, in ogni standard, di un predicato che codifichi l'attribuzione di responsabilità intellettuale.

Attribuire responsabilità a chi ha prodotto le descrizioni va assieme all'esigenza di valorizzare le descrizioni fornendo strumenti di contestualizzazione e fonti di riferimento da cui le informazioni descrittive sono tratte, affiancando il grado di certezza di ogni asserzione fatta.[162]

Valorizzare significa, in questo contesto, usare, ovvero anche mettere a disposizione dell'utente finale, le fonti capaci di aumentare l'esperienza conoscitiva attraverso la dichiarazione di dati (e fonti) attendibili. Ma anche usare modelli ontologici per fornire alle descrizioni quella dimensione concettuale che permetta l'inferenza, la scoperta del non detto e che dia il giusto ruolo a ogni elemento della descrizione, quello che ci consente di dire che una speciale data è la data di digitalizzazione e non quella della creazione dell'esemplare analogico.

Attribuzione di paternità degli asserti in ottica di valorizzazione, ovvero dichiarazione della loro provenienza (o anche *provenance*, per cui si veda par. 3.3) in dimensione spazio-temporale, assieme alle fonti di riferimento diventano le chiavi attraverso cui aumentare la capacità comunicativa dei contenuti delle descrizioni. Ed ecco che il tema della descrizione in un ambiente digitale si inserisce in una più ampia riflessione sui sistemi di trattamento delle informazioni attraverso la lente di osservazione delle DH. In un quadro insomma che pone la descrizione al centro di un ampio tema di analisi, gestione, divulgazione dell'informazione veicolata dai dati.

Questa riflessione sulla valorizzazione è strettamente connessa al concetto di arricchimento.[163] Se, tradizionalmente, con il termine arricchimento, nel contesto dei LOD, si allude alla creazione di un sistema di collegamento a fonti autorevoli a scopo di "riconciliazione"[164] che sia per i luoghi, i concetti o le persone (ad es. il canonico owl:sameAs per collegare un'entità alla sua forma controllata sui vocabolari, come i già citati VIAF, LCSH, Getty, GeoNames ecc., su cui comunque una riflessione andrebbe fatta), con arricchimento vogliamo anche alludere alla necessità di rafforzare il significato dei valori degli elementi della descrizione con una maggiore tipizzazione del loro ruolo (data di creazione dell'esemplare analogico, diversa dalla data di realizzazione della

versione digitale, ancora diversa dalla data di pubblicazione dell'oggetto digitale). E se questo non viene fatto a livello di metadati, è necessario un ulteriore filtro affinché la conversione in LOD sia davvero affidabile.

Ecco che la trasformazione di dati a semantica implicita in LOD andrà demandata a chi sappia leggere i dati della scheda ed eventualmente arricchire l'informazione con ulteriori dati provenienti da fonti certificate, procedendo alla traduzione dei dati di origine in un set di triple attendibili tanto quanto le descrizioni originarie. E così non si correrà il rischio di associazioni errate (la data della scheda associata alla paternità dell'opera) o di filtri svincolati dal contesto (il filtro data mediato dal contesto in cui quella specifica data attesta una dimensione temporale potenzialmente multiforme). E così anche gli asserti saranno affidabili. Ancora più se documentati, come si diceva, attraverso esplicita menzione delle fonti utilizzate per l'arricchimento: dai sistemi di controllo d'autorità, all'*harvesting* (potenzialmente semi-automatico) di relazioni latenti già presenti sul LOD *cloud*.

E assieme all'arricchimento, andrà quindi formalizzato quel livello di rappresentazione stabilito dalla definizione di responsabilità sulle affermazioni fatte in merito agli oggetti di interesse, ovvero un meta-livello necessario per attribuire la paternità delle asserzioni e legare le dichiarazioni fatte su agenti, ruoli ed eventi all'interprete che ha estratto queste informazioni dai documenti. La provenienza degli asserti è fondamentale sia per attribuire la paternità delle asserzioni fatte sui documenti (chi ha detto cosa), sia per individuare le responsabilità della creazione dei dati (chi ha formalizzato in tripla RDF qualcosa asserita da qualcuno).

Con le ontologie infatti riusciamo a risolvere anche questo aspetto di descrizione. PROV-O,165 ad es., è considerato standard per la rappresentazione della provenienza dei dati. Ogni standard di descrizione dovrebbe quindi non rinunciare alla sua espressività, ma provvedere all'aggiunta di un *pattern* per formalizzare il principio di provenienza.

Quando i dati descrittivi degli oggetti del patrimonio culturale siano l'esito della necessità di documentare non solo dati di carattere fattuale, ma dati esito della ricerca scientifica condotta su quei dati, ecco che la nozione di descrizione abbraccia quella di interpretazione.

Ecco perché, con l'obiettivo di portare in concreto un lavoro di ricerca che, partendo da questi presupposti teorici, ha riflettuto sulla costruzione di una concettualizzazione formale, si vuole introdurre brevemente un modello. HiCO,166 è una ontologia nata per descrivere il contesto di oggetti culturali in quanto prodotto del *workflow* interpretativo di un agente. Questo modello consente di gestire affermazioni (nella forma di asserti o triple) autorevoli su informazioni e/o asserzioni contraddittorie. Ogni asserto è infatti concepito come una lettura soggettiva di un interprete su uno specifico livello della fonte da cui tale lettura viene estratta, come ad es. un testo, o meglio l'espressione del testo della fonte. Un testo è per sua natura un oggetto multi-livellare: HiCO, acquisendo il modello FRBR, cerca di far fronte a questa complessità per poter valutare i livelli appropriati alla rappresentazione dei diversi aspetti del processo interpretativo (Fig. 4).

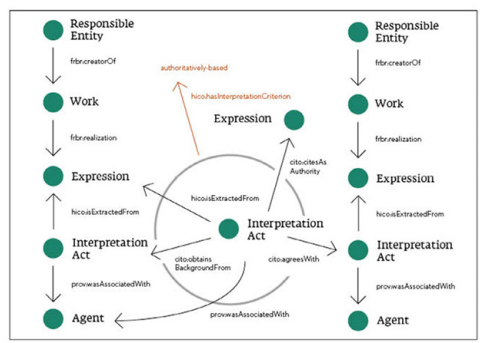

Figura 4 - HiCO. A sinistra un oggetto culturale (ad es. una fonte primaria).
A destra un altro oggetto culturale che a che fare con il primo (ad es. una fonte secondaria, o l'edizione a stampa di quella fonte). Al centro l'atto interpretativo che collega i due oggetti (la fonte e la sua edizione) e crea una nuova espressione (gli asserti in una possibile edizione digitale che usa la fonte primaria e riusa i dati della fonte secondaria). Grafico a cura di Marilena Daquino

Partendo dal problema della provenienza, HiCO estende l'analisi alle entità coinvolte nel processo interpretativo, proponendo un *workflow* dettagliato per stabilire, secondo criteri condivisibili, quali asserzioni possano essere considerate autorevoli e/o meglio documentate rispetto a un'asserzione priva di contesto che ne avvalori il contenuto.

In particolare HiCO – che riusa FRBR, ma anche PRO[167] per la gestione dei ruoli, il *pattern* N-ary Participation[168] per il concetto di partecipazione a un evento e appunto PROV-O per il tema della provenienza dei dati canonicamente intesa – permette di descrivere l'atto interpretativo di un agente sulla base di "tipologie" di interpretazioni e "criteri" interpretativi e, attraverso l'ontologia CiTO,[169] consente di creare relazioni fra interpretazioni riferite allo stesso oggetto culturale fatte da autori/editori/commentatori diversi.

Non si tratta quindi sono di documentare chi ha detto qualcosa secondo la tradizione della *provenance*, ma anche dichiarare chi ha creato le triple (ad es. l'editore digitale), dando il giusto valore agli asserti anche in caso di posizioni contradditorie sullo stesso oggetto, o prevedendo la coesistenza di punti di vista diversi, basati magari su altrettante diverse fonti di riferimento. Ed è questo l'aspetto veramente importante del processo: dare valore alla molteplicità delle interpretazioni che gli studiosi hanno elaborato sullo stesso oggetto d'analisi. Non c'è dato senza l'interpretazione che quel dato ha generato.

Un risultato dell'applicazione di HiCO è, ad es., il progetto mAuth,[170] un'applicazione che, attraverso l'uso del principio della convivenza di atti interpretativi difformi sullo stesso oggetto di analisi, utilizza una serie di parametri per proporre all'utente finale un ragionevole range di autorevolezza delle diverse interpretazioni:

> mAuth is a tool for art historians, data collection managers, and curious, who want to collect information – historians' opinions, motivations, bibliographic references, and images – about the history of authorship attributions related to artworks of the Modern Art (15-16th centuries). It is based on a semantic crawler that harvests authorship attributions in the Web of Data and returns the list of contradictory statements sorted by their authoritativeness.[171]

Attestare la molteplicità delle attribuzioni di paternità alla stessa opera, fornendo nel contempo all'utente una supposizione di autorevolezza delle attribuzioni, basata su parametri stabili di valutazione, è un servizio a valore aggiunto del tema della complessità, in prospettiva di osservazione trasversale dei contenuti culturali.

Sempre con un'ottica orientata al tema della molteplicità, si vuole concludere allora questa sezione con un altro esempio pratico. Partendo dall'esperienza condotta in seno alla collaborazione con la Fondazione Zeri per la valorizzazione del

patrimonio rappresentato dall'archivio fotografico[172] (attraverso la creazione del dataset Zeri[173] del progetto Zeri & LODe),[174] l'arricchimento semantico trova anche applicazioni utili a rielaborare l'atto interpretativo come strumento per favorire l'esperienza conoscitiva dell'utente finale. Perché, come vedremo oltre, l'arricchimento semantico ha senso se e solo se è possibile sviluppare applicazioni in grado di riusare quei dati per aumentare la conoscenza (per cui si veda cap. 5).

Un esempio tratto dal progetto Zeri & LODe potrà giovare (Fig. 5). Questo esempio mostra come sia possibile inferire l'autorevolezza di un'attribuzione (che non significa dare per certa, ma suggerire al lettore come la più accreditata), quando questa si basi sull'esplicitazione di tipo e criterio adottato per l'atto interpretativo.

Arricchimento attraverso fonti certificate che supportino ipotesi interpretative e attribuzione di paternità delle azioni, che dia conto del tema della complessità e del molteplice, sono allora i due aspetti da cui è opportuno muovere per alcune riflessioni conclusive.

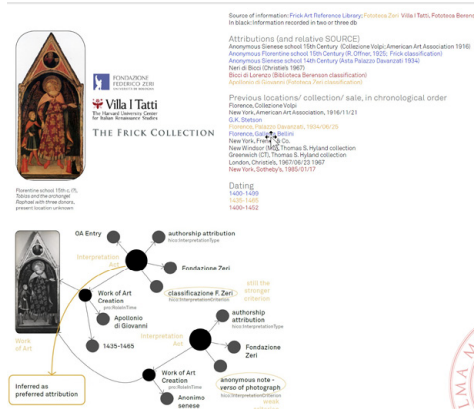

Figura 5 - Si veda, nella parte superiore, l'esempio di un'opera (*Tobia e l'arcangelo Raffaele*) e le diverse istituzioni che l'hanno descritta (Fondazione Zeri, Villa I Tatti e Frick). Subito a fianco è possibile leggere le diverse attribuzioni di paternità, comprensive di fonti di riferimento e datazioni, proposte dalle diverse istituzioni. Nella parte inferiore dell'immagine è possibile capire quali siano i criteri adottati per inferire il motivo dell'individuazione della Fondazione Zeri come l'attribuzione più attendibile, sulla base dei parametri dichiarati: autorevolezza della classificazione Zeri perché associata ad un atto interpretativo [hico:InterpretationAct] dotato di una

tipizzazione [hico:InterpretationType] e
l'esplicitazione di un criterio intepretativo
[hico:InterpretationCriterion], con esplicita menzione
della fonte usata per l'attribuzione di paternità.
Modello a cura di Marilena Daquino. Si ringrazia
Francesca Mambelli per la scheda di esempio

2.4 Le nuove prospettive dell'integrazione dei saperi

Da quanto introdotto, deduciamo che le frontiere della descrizione degli oggetti culturali, in particolare quando siano sul Web, devono estendersi ai nuovi modelli semantici che la rete mette a disposizione e che sono l'esito del lavoro di ricerca condotto sull'espressività di contenuti e contenitori, concetti e supporti materiali.

Usare certe metodologie e strumenti semantici per analizzare artefatti culturali è il mezzo per estendere il patrimonio informativo spesso rappresentato da *silos* di dati. Questa estensione può essere perseguita, come già detto, sia con l'introduzione di nuovi contenuti provenienti da fonti esterne e liberamente riusabili, che con l'estrazione di informazione già disponibile in forma implicita, ovvero non ancora formalizzata in un particolare modello descrittivo, ma magari disponibile sotto forma di dati non strutturati.

Pensiamo alle biblioteche. E proviamo ad andare oltre l'arricchimento determinato dalla sola riconciliazione attraverso record d'autorità. Perché la descrizione catalografica in ottica LOD non deve poter beneficiare dell'arricchimento semantico che il Web stesso può fornire, con lo scopo di valorizzare l'esperienza utente? Riusando magari quanto nel LOD *cloud* è già disponibile? O anche impiegando i *data lake* disponibili su sistemi distribuiti? E allora assieme a documentare fonti, persone, luoghi e concetti, perché non creare relazioni con dati provenienti da altre fonti dati attendibili? Perché non acquisire immagini affidabili relative ai profili delle persone, mappe dei luoghi, concetti in forma di *tag cloud* o ancora perché non collegare, laddove possibile, dati bibliografici a potenziali fondi archivistici, o anche a dati relativi a collezioni museali? O ancora perché non erogare servizi per l'utente finale come la geolocalizzazione o l'accesso a linee temporali navigabili? E viceversa, perché non contribuire affinché una query su un libro, condotta attraverso Google, invece di restituirci un infobox che

estrae dati solo da Wikipedia e da siti commerciali o poco più, non ci restituisca i dati provenienti dal catalogo SBN o da WordCat? E perché non consentire alla molteplicità dei diversi atti interpretativi che nel tempo possono essersi succeduti (proposta di nuove datazioni basate sulla scoperta di nuove fonti; diverse attribuzioni di paternità; nuove localizzazioni di artefatti) di essere documentati? Così come la critica del testo dà conto della storia della tradizione, analogamente i sistemi di metadazione dovrebbero consentire la convivenza di posizioni anche antitetiche sui dati. Magari in una prospettiva che veda nel *crowdsourcing*, e quindi nella collaborazione partecipata alle intraprese di descrizione del patrimonio, una soluzione percorribile.

È evidente che si avrà bisogno di tecnologia, di supporto istituzionale e di professionalità in grado di recepire questa sfida così importante, ma quanto mai essenziale. Ben consapevoli che il passaggio non sarà indolore, esso rappresenta la svolta di un percorso che ripensa il ruolo delle fonti nell'ottica di una comunicazione finalizzata a migliorare l'esperienza utente. Archivi, biblioteche e musei devono ripensare ai LOD in una prospettiva nuova, che può beneficiare dell'esperienza delle DH nel contesto di un Web autenticamente semantico.

Spetta quindi alle istituzioni culturali riprendere in mano quel ruolo di mediatori del sapere che a oggi sta passando nelle mani delle imprese commerciali. E questa è la sfida del rapporto fra LIS e DH: valorizzare la descrizione degli oggetti culturali attraverso la lente dell'interpretazione, che con la creazione, la selezione e l'utilizzo dei modelli ontologici più adeguati a finalità e obiettivi di ricerca, possa arricchire i dati culturali, e quindi l'esperienza informativa, attraverso l'integrazione di risorse eterogenee e trasversali, così come lo sono, per loro stessa natura, gli artefatti culturali conservati nelle biblioteche, negli archivi e nei musei, ma anche i progetti digitali che su quei dati sono basati.

Non andrà allora dimenticato un aspetto che arricchisce la discussione critica su questi temi: la filologia, ovvero in particolare la critica del testo. Lavorare su contenuti culturali è un'esperienza multidimensionale: se da un lato attribuisce al concetto di dato la dimensione paratestuale necessaria a dare contesto a un artefatto culturale, dall'altro, quando quell'artefatto sia una fonte testuale come un documento d'archivio o un manoscritto (ma anche un incunabolo o una cinquecentina), il sistema di metadatazione passa

dalla dimensione del paratesto a quella del testo. Ecco che i LOD allora, oltre a essere lo strumento necessario ad arricchire il corredo paratestuale delle risorse culturali, diventano lo strumento per esplicitare la semantica che implicitamente è veicolata dalla sequenza di stringhe di caratteri di un testo digitale, ovvero da dati non strutturati.

Quando si adotti un approccio consapevole del rapporto testo/paratesto allora dati e metadati di natura eterogenea (archivistici, filologico/testuali, bibliografici e museali), assieme alla computer science e alla mediazione delle DH, potranno configurarsi come il nuovo strumento di rappresentazione della conoscenza in un *knowledge graph* delle edizioni scientifiche davvero espressivo e utile anche per la ricerca. Non si tratterà solo di annotare il testo, ma di ricostruire la storia di quel testo e dar conto, in potenza, della complessità della tradizione, dando spazio ai diversi punti di vista, o anche ai diversi atti interpretativi, che si sono succeduti nel tempo e che hanno una diversa paternità (basti pensare a quanta informazione può trasmettere l'apparato di un'edizione critica).

Si vogliono allora qui menzionare due recenti progetti che raccontano quest'ottica e allo stesso tempo ci forniscono alcuni presupposti teorici che affronteremo nei prossimi capitoli, ovvero l'edizione digitale, o anche il *knowledge site*, delle lettere di Vespasiano da Bisticci[175] e l'edizione digitale semantica del quaderno di appunti di Paolo Bufalini.[176]

In entrambi gli esperimenti, ultima release 2020, la filologia è stata il cuore dell'approccio critico alle fonti primarie, rappresentate da documenti archivistici (lettere, in vario supporto e provenienza, e un quaderno di appunti manoscritti) assieme a risorse bibliografiche manoscritte (codici miscellanei e codici di dedica). L'edizione digitale ha significato una trascrizione diplomatica dei testimoni, una marcatura XML/TEI sulla base del modello interpretativo elaborato e una trasformazione in LOD, con una semantizzazione determinata dalla scelta e dall'adozione delle più appropriate ontologie, capaci di formalizzare l'atto ermeneutico e tradurre in conoscenza esplicita e computabile l'informazione fornita attraverso il markup. Tanto l'edizione si è giovata dei vocabolari controllati elaborati nei contesti archivistici e librari per qualificare persone, luoghi e soggetti/concetti o categorie concettuali in generale (riusando i già citati LC authorities, Getty vocabularies, VIAF, DDC e UDC ecc.), quanto la descrizione

archivistica e bibliografica delle fonti potrebbe giovarsi del sapere che l'accesso al full-text delle fonti può portare alla metadazione descrittiva. E allora solo un mutuo e reciproco puntamento, che il sistema delle triple consente, può arricchire il *knowledge graph* del patrimonio culturale.

Con una precisazione, che, di nuovo, la teoria a fondamento della filologia, con il suo approccio testo-centrico, ci insegna. Non si può, in linea di principio, standardizzare la semantica. Ragioniamo per un attimo sul testo in senso filologico, adottando al contempo un approccio linguistico, ovvero acquisiamo il principio dell'analisi del testo, l'esplorazione del suo contenuto, ovvero la capacità di riconoscere il significato delle occorrenze. Non immaginiamo di lavorare su un testo in particolare, ma ragioniamo su un corpus globale, quale potrebbe essere l'intero Web, con i suoi dati. Se consideriamo un termine lessicale, dobbiamo riconoscere che non esiste un unico *knowledge graph*, costituito dalle relazioni con altri termini (e dalle triple che le rappresentano), perché l'insieme di queste relazioni non è dato a priori, ma dipende dai diversi contesti in cui il termine viene usato. Si dovrà allora riflettere sul rapporto tra pura metadatazione, o descrizione dell'oggetto culturale, e l'analisi del suo contenuto. Lavorare sul testo pieno consente di dare il giusto contesto all'uso del vocabolario e restituire la sfaccettatura prismatica dell'atto interpretativo. Ricordiamo che abbiamo a che fare con oggetti che sono sempre complessi, interconnessi e multi-dimensionali. Se la variabilità semantica è determinata dal contesto, così come le concordanze ci insegnano (valutare il significato delle occorrenze a seconda della posizione delle parole nel testo), non è possibile non pensare alle variabili temporali (valutare il cambiamento del significato dei lemmi nel tempo), culturali (valutare il significato delle parole nelle diverse culture) e individuali (valutare la soggettività dell'interprete e il suo bagaglio esperienziale), che naturalmente spostano il punto di osservazione del grafo e il significato dei termini nella loro rete relazionale. Ma questi grafi della conoscenza, che documentano il contenuto di ogni risorsa, andranno posti fra di loro in relazione per raccontare un percorso, individuare costanti o valorizzare differenze.

In definitiva: "include links to other URIs so that they [sottointeso 'people'] can discover more things." dice Berners-Lee.[177] Ed è solo in quest'ottica che si potrà parlare di vero avanzamento della conoscenza. Perché è il collegamento che

acquisisce, e allo stesso tempo stabilisce, il contesto, arricchendo l'esperienza di navigazione. Detto in altri termini, per arrivare a un 5 stelle LOD: "link your data to other people's data to provide context".178 Ma il contesto varia al variare dell'asserto, e analogamente le entità assumono un valore e un'identità in quanto collocate in un contesto in cui sono gli elementi in relazione che ne determinano il significato. Ed è dalle relazioni con altri dati che si crea la conoscenza. Quindi il contesto diventa l'elemento determinante per attribuire significato e valorizzare le possibili tipologie di relazioni.

Fornire contesto dunque. Il tema che, a partire da questa riflessione su LIS e DH, rappresenta il cuore dell'analisi che si intende portare avanti nel prossimo capitolo.

83 Paolo Bisogno, *Teoria della documentazione*, Milano, Franco Angeli, 1979, p. 18.
84 Harold Borko, *Information science: What is it?*, cit., p. 3. Grassetto nostro.
85 Si veda, ad es., Arlene G. Taylor, *The organization of information*, Englewood, Libraries Unlimited, 1999.
86 Una prospettiva di insieme nell'ultima edizione dell'enciclopedia sulla LIS, *Encyclopedia of Library and Information Sciences*, edited by John D. McDonald, Michael Levine-Clark, Fourth edition Vol. 1-7, Boca Raton, FL, CRC Press, 2017. Di particolare interesse è il progetto, avviato nel 2018, per la realizzazione di una versione elettronica di questo importante progetto enciclopedico, https://www.librarianshipstudies.com/2018/03/library-information-science-encyclopedia.html.
87 Si veda, ad es., il punto di vista espresso da Birger Hjørland, *Information Science and Its Core Concepts: Levels of Disagreement*, in *Theories of Information, Communication and Knowledge*, edited by Fidelia Ibekwe-SanJuan, Thomas Dousa, Studies in History and Philosophy of Science 34, Dordrecht, Springer, 2014, p. 205-235, https://doi.org/10.1007/978-94-007-6973-1_9.
88 Molto interessante la posizione di Vivarelli che, al di là dell'analisi dei punti di contatto, auspica una più profonda collaborazione fra le discipline, Maurizio Vivarelli, *Digital humanities e culture documentarie: un modello di analisi, valutazione, interpretazione*, "AIB studi", 60 (2020), 3, p. 553-589, https://doi.org/10.2426/aibstudi-12471.
89 Una disamina completa in Birger Hjørland, *Library and information science: practice, theory, and philosophical basis* "Information Processing and

Management", 36 (2000), p. 501-531.
90 Un *survey* in Ying Zhang, Shu Liu, Emilee Mathews, *Convergence of digital humanities and digital libraries*, "Library management", 36 (2015), 4-5, p. 362- 367, https://doi.org/10.1108/LM-09-2014-0116.
91 Chris Alen Sula, *Digital Humanities and Libraries: A Conceptual Model*, "Journal of Library Administration", 53 (2013), 1, p. 10-26, https://doi.org/10.1080/01930826.2013.756680.
92 Una semplice query condotta su LISTA è abbastanza eloquente: dai 783 risultati della ricerca "digital humanities" condotta sugli abstracts delle pubblicazioni dal 1998 al 2012, si passa a 708 nel solo periodo 2012-2020.
93 Chris Alen Sula, *Digital Humanities and Libraries: A Conceptual Model*, cit., p. 13.
94 Questa classificazione è riportata nel post di Ben Showers, *Does the library have a role to play in digital humanities?*, "JISC", 2012, http://infteam.jiscinvolve.org/wp/2012/02/23/does-the-library-have-a-role-to-play-in-the-digital-humanities.
95 A cui aggiungiamo l'importante tema dell'*information retrieval*, a riconferma dell'importanza dei sistemi di analisi del testo nella LIS, per cui si veda Maurizio Lana, *Digital humanities e biblioteche*, cit., p. 199-200.
96 Marilena Daquino, Francesca Tomasi, *Digital Humanities e Library and Information Science. Through the lens of knowledge organization*, "Bibliothecae.it", 5 (2016), 1, p. 130-150, https://doi.org/10.6092/issn.2283-9364/6109.
97 Sul tema della molteplicità dei punti di vista, anche contradditori, che si possono esprimere sui dati culturali si veda: Gioele Barabucci, Francesca Tomasi, Fabio Vitali, *Supporting Complexity and Conjectures in Cultural Heritage Descriptions*, in *Collect and Connect: Archives and Collections in a Digital Age 2020 (COLCO 2020)*, "CEUR Workshop Proceedings", 2810 (2021), p. 104-115, http://ceur-ws.org/Vol-2810/paper9.pdf.
98 A questo proposito si veda Helena Robinson, *Remembering things differently: Museums, libraries and archives as memory institutions and the implications for convergence*, "Museum Management and Curatorship", 27 (2012), 4, p. 413-429.
99 Istituto Centrale per il Catalogo Unico (ICCU), https://www.iccu.sbn.it/it.
100 Istituto Centrale per gli Archivi (ICAR), https://www.icar.beniculturali.it.
101 Istituto Centrale per il Catalogo e la Documentazione (ICCD), http://www.iccd.beniculturali.it.
102 International Federation of Library Associations and Institutions (IFLA), https://www.ifla.org.
103 International Council on Archives (ICA), https://www.ica.org.
104 International Council of Museums (ICOM), https://icom.museum.
105 *International Standard for Bibliographic Description* (ISBD), Consolidated

Edition 2011; trad. it., Roma, ICCU, 2012, https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/publications/series/44-it.pdf.

106 *Anglo-American Cataloguing Rules* (AACR). "First published in 1967 and edited by C. Sumner Spalding, a second edition (AACR2) edited by Michael Gorman and Paul W. Winkler was issued in 1978, with subsequent revisions (AACR2R) appearing in 1988 and 1998; all updates ceased in 2005", https://en.wikipedia.org/wiki/Anglo-American_Cataloguing_Rules.

107 *Resource Description and Access* (RDA), trad. it., Roma, ICCU, 2015, https://www.iccu.sbn.it/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf. Toolkit, https://www.rdatoolkit.org/. Si veda in particolare Carlo Bianchini, Mauro Guerrini, *Introduzione a RDA. Linee guida per rappresentare e scoprire le risorse*, Milano, Editrice Bibliografica, 2014.

108 *Functional Requirements for Authority Data* (FRAD), final report IFLA 2008, rev. 2013, https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frad/frad_2013.pdf.

109 *Functional Requirements for Subject Authority Data* (FRSAD), final report IFLA 2010, https://www.ifla.org/wp-content/uploads/2019/05/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf.

110 Pat Riva, Patrick Le Bœuf, Maja Žumer, *IFLA Library Reference Model* (LRM), 2017, trad. it., Roma, ICCU, 2020, https://repository.ifla.org/bitstream/123456789/44/1/ifla-lrm-august-2017_rev201712-it.pdf.

111 BIBFRAME, https://www.loc.gov/bibframe, version 2, 2016: "BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world that is grounded in Linked Data techniques". L'ontologia si può consultare all'indirizzo https://id.loc.gov/ontologies/bibframe.html.

112 Per una visione d'insieme su standard di catalogazione e vocabolari controllati nel campo delle biblioteche si può leggere l'attività del W3C incubator group su Library Linked Data (lld) e in particolare il report del 2011 *Datasets, Value vocabularies and Metadata Element Sets*, https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/.
Oppure si può consultare il sito dell'IFLA, nella sezione dedicata agli standard, https://www.ifla.org/standards, dove però manca l'attività della Library of Congress, consultabile dalle pagine https://www.loc.gov/librarians/standards.

113 Dewey Decimal Classification (DDC), versione digitale scaricabile da OCLC, https://www.oclc.org/en/dewey/webdewey.html.

114 Universal Decimal Classification (UDC), https://udcc.org.

115 Colon Classification (CC), un outline con alcuni esempi dal sito ISKO, http://www.iskoi.org/doc/colon.htm.

116 Library of Congress Subject Headings (LCSH), https://id.loc.gov/authorities/subjects.html. Anche per i sistemi di classificazione e

soggettazione si veda il report appena menzionato del W3C su *Datasets, Value vocabularies and Metadata Element Sets*, cit.

117 Nuovo soggettario, a cura della Biblioteca nazionale centrale di Firenze, https://thes.bncf.firenze.sbn.it.

118 Come mostrano, a titolo di esempio, l'esperimento italiano di Manus OnLine dell'ICCU, https://manus.iccu.sbn.it, o l'internazionale Fragmentarium, https://fragmentarium.ms.

119 International Image Interoperability Framework (IIIF), sito ufficiale http://iiif.io. IIIF (comunemente pronunciato "triple-i f") è un framework pensato per fornire un servizio ad applicativi software finalizzati a visualizzare, ma anche annotare, collezioni di dati. Nato e sviluppatosi in ambito biblioteconomico, l'IIIF si basa sulla definizione di interfacce di programmazione (API) che i programmi, sia lato client che server, possono implementare, per permettere così l'interscambio delle immagini. Sviluppato e supportato da istituzioni come la Bibliothèque Nationale de France, la British Library e la Bodleian Library, insieme a centri di ricerca ben noti nel mondo delle biblioteche digitali, tra cui la Cornell University, l'IIIF sta diventando uno standard de facto, così come accaduto con la TEI. Per un'introduzione su IIIF si veda Alberto Salarelli, *International Image Interoperability Framework (IIIF): a panoramic view*, "JLIS", 8 (2017), 1, p. 50-66, https://dx.doi.org/10.4403/jlis.it-12090.

120 "JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale", https://json-ld.org.

121 *International Standard for Archival Authority Records for Corporate bodies, Persons, Families* (ISAAR-CPF), first edition 1996, ora siamo alla seconda edizione 2003, accessibile dal sito ICA in varie lingue, https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd; trad. it. a cura di Stefano Vitali, in "Rassegna degli Archivi di Stato", 63 (2003), 1, p. 191-333.

122 Encoded Archival Context (Corporate bodies, Persons, Families) (EAC-CPF), https://eac.staatsbibliothek-berlin.de.

123 *International Standard for Describing Functions* (ISDF), first edition 2007, dal sito ICA in varie lingue, https://www.ica.org/en/isdf-international-standard-describing-functions.

124 *International Standard for Describing Institutions with Archival Holdings* (ISDIAH), first edition 2008, dal sito ICA in varie lingue, https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings.

125 Records in Contexts – Conceptual Model (RiC-CM), version 0.1, 2016, version (official) 0.2, 2021, https://www.ica.org/en/records-in-contexts-conceptual-model.

126 Records in Contexts – Ontology (RiC-O), https://www.ica.org/

standards/RiC/ontology.html.
127 Ontology for Archival Description (OAD), dal progetto Reload, https://labs.regesta.com/progettoReload/oad-ontology.
128 Encoded Archival Context (Corporate bodies, Persons, Families) Ontology, http://culturalis.org/eac-cpf.
129 Maria Letizia Mancinelli, *Gli standard catalografici dell'Istituto Centrale per il Catalogo e la Documentazione,* in Roberta Tucci, *Le voci, le opere e le cose. La catalogazione dei beni culturali demoetnoantropologici*, Roma, ICCU, 2018, p. 279-302. Si veda anche l'elenco degli standard dal *repository* su Github, https://github.com/ICCD-MiBACT/Standard-catalografici.
130 CIDOC Conceptual Reference Model (CRM), https://www.cidoc-crm.org. Latest version, https://cidoc-crm.org/version/version-7.1.1.
131 Sull'importanza di CIDOC CRM si veda in particolare Martin Doerr, *Ontologies for Cultural Heritage*, in *Handbook on Ontologies*, edited by Steffen Staab, Rudi Studer, Berlin, Springer, 2009, p. 463-486.
132 Dublin Core (DC), https://www.dublincore.org.
133 DC Terms, https://www.dublincore.org/specifications/dublin-core/dcmi-terms.
134 Friend Of A Friend (FOAF), http://xmlns.com/foaf/spec.
135 https://schema.org.
136 Una panoramica completa degli standard nel pamphlet *Seeing Standards: A Visualization of the Metadata Universe*, a cura di Jenn Riley, 2009-2010, http://jennriley.com/metadatamap.
137 Metadati Amministrativi e Gestionali (MAG), https://www.iccu.sbn.it/export/sites/iccu/documenti/manuale.html.
138 Metadata Encoding and Transmission Standard (METS), http://www.loc.gov/standards/mets.
139 Preservation Metadata (PREMIS), https://www.loc.gov/standards/premis.
140 Una buona visione d'insieme sui sistemi di classificazione dei metadati in Jenn Riley, *Understanding metadata. What is metadata and what is it for?*, a Primer Publication of the National Information Standards Organization (NISO), USA, 2017, https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf.
141 Si veda ancora il pamphlet *Seeing Standards: A Visualization of the Metadata Universe*, cit.
142 Si veda il già citato report del W3C *Datasets, Value vocabularies and Metadata Element Sets*, cit.
143 Wordnet, https://wordnet.princeton.edu.
144 Virtual International Authority File (VIAF), https://viaf.org.
145 Linked Data Service, https://id.loc.gov.
146 Getty Vocabularies, http://www.getty.edu/research/tools/vocabularies. Si tratta dei thesauri Art & Architecture Thesaurus (AAT), Cultural Objects Name Authority (CONA), Getty Thesaurus of Geographic Names (TGN),

Union List of Artist Names (ULAN), Getty Iconography Authority (IA) e delle regole di catalogazione Categories for the Description of Works of Art (CDWA).

147 GeoNames, http://www.geonames.org.

148 Pleiades, https://pleiades.stoa.org.

149 Dati del Sistema Archivistico Nazionale (SAN), http://dati.san.beniculturali.it.

150 CulturaItalia, http://www.culturaitalia.it.

151 Internet Culturale, https://www.internetculturale.it.

152 Alphabetica, https://alphabetica.it/. Per una descrizione del progetto si veda Simonetta Buttò, *Alphabetica, il nuovo portale per la ricerca integrata: un salto di qualità per le biblioteche italiane*, "DigItalia. Rivista del digitale nei beni culturali", 14 (2019), 1, p. 9-28, https://doi.org/10.36181/digitalia-00010.

153 Catalogo generale dei Beni Culturali, https://www.catalogo.beniculturali.it.

154 Dati in LOD, https://dati.beniculturali.it.

155 Varrà la pena ricordare che i dati della pubblica amministrazione sono rilasciati in formato aperto e sono accessibili dalla piattaforma dati.gov.it. Sicuramente utile anche il lavoro condotto per la creazione delle linee guida per la creazione di dati aperti. Agenzia per l'Italia Digitale (AGID), *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, ultima release 2020, https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/index.html.

156 Chryssoula Bekiari, Martin Doerr, Patrick Le Bœuf, Pat Riva, *Definition of FRBROO. A Conceptual Model for Bibliographic Information in Object-Oriented Formalism*, IFLA 2016, http://repository.ifla.org/bitstream/123456789/659/1/frbroo_v_2.4.pdf.

157 Su questo tema si veda Gioele Barabucci, Francesca Tomasi, Fabio Vitali, *Modeling data complexity in public history and cultural heritage*, in *Handbook Digital Public History*, edited by Serge Noiret, Mark Tebeau, Gerben Zaagsma, Berlin, De Gruyter, 2022 (in corso di stampa).

158 Per cui si veda, ad es., Silvia Bruni et al., *Verso l'integrazione tra archivi, biblioteche e musei. Alcune riflessioni*, "JLIS.it", 7 (2016), 1, p. 225-44, https://doi.org/10.4403/jlis.it-11482.

159 Europeana Data Model (EDM), https://pro.europeana.eu/page/edm-documentation.

160 Si veda in particolare EDM, *Mapping Guidelines*, version 2.4, 2017, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf.

161 Significativo che la stessa Agenzia per l'Italia Digitale (AGID) rifletta su questo tema, per cui si veda Commissione di Coordinamento SPC, *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, versione

2.0, 2012, https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf.

162 Sul tema dell'importanza di documentare il grado di certezza delle affermazioni si veda Michael Piotrowski, *Accepting and Modeling Uncertainty*, "Zeitschrift für Digitale Geisteswissenschaften", 2019, http://doi.org/10.17175/sb004_006a.

163 Sul tema dell'arricchimento semantico dei dati si veda in particolare Marcia Lei Zeng, *Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article*, "El profesional de la información", 28 (2019), 1, https://doi.org/10.3145/epi.2019.ene.03. Linee guida cui riferirsi per il tema dell'arricchimento, sono quelle riportate nel report del gruppo di Europeana, Antoine Isaac, Hugo Manguinhas, Juliane Stiller, Valentine Charles, *Report on enrichment and evaluation*, The Hague, Netherlands, Europeana Task Force on Enrichment and Evaluation, 2015, https://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf.

164 Spesso vengono utilizzati tool per agevolare questa operazione. Si veda, ad es., OpenRefine, https://openrefine.org.

165 The Provenance Ontology (PROV-O), http://www.w3.org/TR/prov-o.

166 Historical Context Ontology (HiCO), http://purl.org/emmedi/hico.

167 Publishing Roles Ontology (PRO), http://purl.org/spar/pro.

168 N-ary Participation, http://ontologydesignpatterns.org/wiki/Submissions:Nary_Participation.

169 The Citation Typing Ontology (CiTO), http://purl.org/spar/cito.

170 Mining Authoritativeness in Art History (Mauth), http://purl.org/emmedi/mauth/search.

171 Marilena Daquino, *Mining Authoritativeness in Art Historical Photo Archives: Semantic Web Applications for Connoisseurship*, Studies on the Semantic Web 40, Berlin, IOS Press, 2019.

172 L'archivio Zeri (conservato presso la Fondazione Zeri dell'Università di Bologna) raccoglie una serie di oggetti culturali complessi: 290.000 fotografie di opere d'arte; 46.000 volumi; 37.000 cataloghi d'asta; 15.000 documenti archivistici.

173 Si veda il sito ufficiale all'indirizzo http://data.fondazionezeri.unibo.it. Da queste pagine è possibile: leggere la descrizione del progetto, acquisire informazioni sui dati (strumenti di accesso e di browsing, licenze, struttura degli URI, standard W3C usati), consultare ontologie, vocabolari, contratti e authorities impiegati e, infine, accedere alla documentazione di progetto, comprese le pubblicazioni.

174 Si veda, fra gli altri, Marilena Daquino et al., *Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data*, "ACM Journal on Computing and Cultural Heritage"

(JOCCH), 10 (2017), 4, http://dx.doi.org/10.1145/3051487.

175 Prima versione *Vespasiano da Bisticci, Lettere*, edizione a cura di Francesca Tomasi, Bologna, AlmaDL - Università di Bologna, 2013, https://doi.org/10.6092/unibo/vespasianodabisticciletters; alla versione integralmente a base semantica si può accedere da *Vespasiano da Bisticci, Lettere*, Knowledge Site, versione 3.0, Bologna, Digital Humanities Advanced Research Centre (/DH.arc), Università di Bologna, 2020, http://projects.dharc.unibo.it/vespasiano.

176 L'edizione si può consultare in *Paolo Bufalini, Appunti (1981-1991)*, a cura di Marilena Daquino, Martina Dello Buono, Francesca Giovannetti, Francesca Tomasi, Bologna, Digital Humanities Advanced Research Centre (/DH.arc), Università di Bologna, 2020, https://doi.org/10.6092/unibo/amsacta/6415.

177 Tim Berners Lee, *Linked Data*, cit.

178 *Ivi*.

3. CONTESTI. OVVERO LA LEZIONE DEGLI ARCHIVI

A proposito di vector semantics:[179]
"The meaning of a word is
its use in the language"
(Wittgenstein, 1953, PI 43)

3.1 Perché i contesti?

Non c'è disciplina che non trovi nella nozione di contesto la chiave di lettura e di interpretazione dei dati più appropriata per rispondere ai propri quesiti di ricerca. Il contesto è cioè quella serie di componenti che attribuiscono significato al dato, o anche che lo arricchiscono. Un dato privo di contesto è un dato non altrimenti interpretabile, tanto sul piano computazionale quanto su quello umanistico, nelle sue declinazioni che qui ci interessano, come quelle linguistica, letteraria, biblioteconomica e archivistica.

Attribuire contesto è l'azione che tipicamente, dal punto di vista computazionale, consente ai dati di diventare informazione e all'informazione di diventare conoscenza.[180] Affinché questa piramide trovi sistematizzazione è opportuno che a ogni passaggio di questo processo tripartito avvenga un'azione computazionale capace di agire sulla trasformazione.

I dati sono entità atomiche, riconosciute dal calcolatore perché rappresentate in un sistema formale, il linguaggio binario, che traduce le azioni in stringhe. Affinché sia possibile estrarre informazioni, i dati devono essere raccolti e rappresentati in base a regole e organizzati in strutture. La relazione che sussiste fra i due concetti di dato e informazione è concetto chiave; diremo che le informazioni vengono rappresentate, cioè comunicate alla

macchina, attraverso i dati. Assumendo l'ottica del pensiero computazionale (*computational thinking*), se l'informazione può essere definita come la notizia relativa a un fatto, cioè il contenuto del dato, il dato è invece l'elemento dell'informazione costituito da simboli che devono essere elaborati e interpretati per fornire informazioni. Diremo quindi che il dato diventa informazione quando è inserito in un contesto tale da essere identificato rispetto a un ambito di riferimento, cioè gli viene associata una struttura.

Computazionalmente allora il contesto è lo strumento che trasforma il dato in informazione, perché attribuisce struttura. E quando abbiamo a che fare con l'informazione? Possiamo dire che l'informazione è un dato dotato di struttura. Può essere un dato semi-strutturato, come quello prodotto quando si utilizzi un linguaggio di markup dichiarativo come XML, o un dato strutturato, come quello inserito, ad es., in un database. Il modello dei dati scelto quindi (gerarchico, reticolare, relazionale o a oggetti) stabilisce il contesto. Diremo che il contesto è quello che determina le possibili tipologie di azioni che sul piano computazionale sarà possibile eseguire sui dati, ovvero è la struttura dati. La capacità dunque di acquisire informazione dal dato è determinata dal contesto.

Se prendiamo una situazione come:

```
<corso xml:id="identificativo">
  <denominazione>Lorem
ipsum</denominazione>
  <docente>
<nome>Francesca</nome>
<cognome>Tomasi</cognome>
  </docente>
</corso>
```

Ne deduciamo che il significato del dato (stringa di caratteri) cambia a seconda del contesto. Nell'esempio di sinistra il contesto è determinato dal markup e il dato diventa informazione attraverso l'etichetta che lo descrive (modello gerarchico). A destra il contesto è la tabella e il dato diventa informazione perché è il valore di un attributo (modello relazionale). E sul piano della semantica implicita diremo che la stessa stringa "Francesca Tomasi" assume una funzione, in questo esempio una connotazione che equivale a un ruolo, differente a seconda del contesto (nel modello gerarchico

è un docente, nel modello relazionale è un autore).

Se sul piano computazionale il contesto è capace di determinare il significato attraverso la struttura dati (l'albero e la tabella nell'esempio precedente), non dissimile è il ruolo del contesto in ambito più strettamente umanistico. Giusto per fare qualche esempio, andiamo dalla teoria della comunicazione di derivazione strutturalista, in cui il contesto è rappresentato dalle circostanze della trasmissione di un messaggio,[181] all'ambito più strettamente linguistico in cui è il sistema integrato di semantica e sintassi, ovvero il contesto d'uso del lessico, che determina il significato delle occorrenze, spostandoci poi sul piano più ampiamente letterario, in cui il contesto è determinato dalle dimensioni storico-culturali, o anche spaziali e temporali, ovvero anche cronotopiche, che stabiliscono la capacità di attribuire significato al testo.

Senza voler scendere ulteriormente nel dettaglio, sicuramente una disciplina che, nella nostra prospettiva critica, ha elaborato chiaramente il concetto di contesto è l'archivistica. Ed è la nozione che adotta questa disciplina che permette alle DH di ampliare la prospettiva critica di analisi, nella duplice direzione computazionale e umanistica in senso lato.

La nozione di contesto in archivistica tipicamente coinvolge una serie di questioni, che sono anche i fattori dell'argomentazione che si intende sostenere nelle prossime pagine; contesto in archivistica significa infatti: soggetto produttore o *creator*, rispetto del principio di provenienza o *provenance*,[182] posizionamento dell'entità nella gerarchia documentaria.[183] Sono questi i concetti che, alla luce delle DH, si estendono e al contempo prevedono l'assunzione di nuovi punti di vista sulla rappresentazione, gestione e organizzazione della conoscenza.

Riconoscere l'importanza del contesto è dunque un'azione fondamentale. Ma date queste premesse è evidente che sarà limitativo parlare di contesto al singolare, perché inevitabilmente ogni processo di rappresentazione dell'informazione richiederà più contesti diversi necessari a stabilire il significato. E consentirci di fare un passo in più: passare dall'informazione alla conoscenza. Perché sono i diversi contesti attraverso cui è possibile leggere l'informazione che permettono l'acquisizione di sapere.

Ma partiamo dai tre principi cardine dell'archivistica: soggetto produttore, *provenance* e posizionamento (dall'albero al grafo).

3.2 Soggetto produttore. La funzione del *creator*

Le scienze del libro e del documento da sempre riconoscono nella funzione del *creator* un ruolo importante nei meccanismi di descrizione delle fonti, tanto bibliografiche quanto archivistiche. Non è un caso se così tanto interesse ha rivestito il principio del controllo di autorità,[184] ovvero della definizione di buone pratiche per qualificare la responsabilità intellettuale del *creator*, o detto in altro modo di quel particolare tipo di *agent* a cui attribuire il ruolo di soggetto responsabile di un qualche rapporto con la fonte.

In archivistica in particolare si parla di soggetto produttore come "l'ente, la famiglia o la persona che ha posto in essere, accumulato e/o conservato la documentazione nello svolgimento della propria attività personale o istituzionale".[185]

Se nella pratica di origine biblioteconomica particolare attenzione viene data al controllo di autorità, con un'enfasi sulla responsabilità intellettuale esercitata dal *creator* in qualità di produttore del contenuto dell'oggetto libro, in ambito archivistico l'esigenza è di documentare il più possibile questa figura come l'entità (nella forma persona, famiglia o ente) che non solo ha prodotto la documentazione, ma anche che l'ha accumulata nel tempo, l'ha conservata e l'ha usata.[186] Quella stessa documentazione testimonia un'azione riconoscibile dal punto di vista della funzione svolta dall'entità stessa, ovvero anche dalla sua attività.

Ecco che gli elementi informativi di contesto, cosi come elaborati nella pratica archivistica, assumono un ruolo fondamentale nell'identificazione del soggetto produttore (supportato dal principio della descrizione separata) e allargano la prospettiva di analisi rispetto alle sole linee guida per la forma normalizzata del nome.

E l'arricchimento della descrizione del *creator* non avviene solo attraverso una profilazione storico-biografica, ma, come ci insegna lo standard ISAAR(CPF), insistendo particolarmente sul principio delle relazioni: relazioni fra soggetti produttori, possibilmente tipizzate in senso gerarchico, familiare o associativo e collocate in un arco temporale; relazioni fra soggetto produttore e complesso documentario prodotto; relazioni fra soggetto produttore e documentazione ascrivibile a un qualche livello (ad es. di natura bibliografica, storico-artistica, museale).

Nel contesto del Web semantico, l'approccio data centrico tipico della descrizione archivistica, assieme all'esigenza di documentare il principio delle relazioni, rende i LOD lo strumento di rappresentazione più adeguato. La nozione di relazione ben si sposa infatti con il meccanismo di funzionamento dei LOD, o meglio con il modello di descrizione RDF, e la tipizzazione della relazione consente di ragionare sui predicati (vocabolari RDF o ontologie) necessari a istituire il collegamento fra entità.

Le profonde connessioni che il soggetto produttore intrattiene con altri soggetti produttori, o in generale con altre entità, è un elemento importante della descrizione archivistica, perché consente di tracciare una fitta rete di rapporti trasversali, tipica di un approccio per triple, e permette di lavorare sulla tipizzazione delle relazioni personali, a un grado di analiticità profonda, determinato dal modello ontologico scelto.[187] Si potrà quindi prevedere di andare oltre una generica relazione associativa, tentando di semantizzare il tipo di associazione (che non necessariamente sarà solo familiare o gerarchica).

Analoga riflessione può essere condotta sulla relazione fra il soggetto produttore e ciò che è ascrivibile alla sua produzione, non solo quindi il complesso documentario – che a sua volta può diventare il soggetto di una nuova tripla – ma anche altre risorse in qualche modo correlate, e magari già altrove descritte e reperibili sul *cloud*. E la relazione fra soggetto e oggetto si potrà giovare dell'analiticità del ruolo (e la sua variabilità nel tempo) che il soggetto può assumere nei confronti dell'oggetto.

Spostandoci sul piano degli standard, non è un caso dunque se lo Schema XML elaborato sulla base dello standard metodologico ISAAR(CPF), che ha rilasciato le linee guida per la costruzione dei record d'autorità, è stato denominato *Encoded Archival Context*, a indicare quanto il soggetto produttore sia il primo strumento per attribuire contesto al complesso documentario ovvero anche a dire che l'entità CPF stabilisce il contesto di produzione della documentazione archivistica.

Nello Schema EAC-CPF, sulla base delle specifiche di ISAAR, e anche nella EAC-CPF ontology,[188] ci sono vari attributi dunque che qualificano il soggetto produttore. In particolare, già ISAAR(CPF) teorizza il principio del *creator* stabilendo, oltre ai canonici elementi obbligatori (tipologia, forma autorizzata del nome, date di esistenza e codice identificativo del record di autorità), una nutrita serie di

attributi nell'area della descrizione (ulteriore set di informazioni che forniscono il contesto dell'entità descritta). Luoghi collegati in qualche modo all'entità, funzioni, occupazioni e attività svolte, assieme alla storia, sono quelli che in particolare ci interessano. Per una serie di ragioni.

Se la forma controllata del nome è quella che ci garantisce l'identificazione univoca, rafforzata dal codice del record di autorità, i luoghi consentono di ragionare sul principio della definizione delle opportune coordinate spaziali, permettendoci di stabilire le connessioni fra l'entità persona, famiglia o ente e le sue localizzazioni. Ovviamente il luogo va immaginato in riferimento al tempo, ovvero collegando lo spazio alla dimensione temporale e alla tipizzazione della relazione che lega l'entità CPF al luogo (entità legata a un luogo in un certo arco temporale). Analogo discorso vale per funzioni, occupazioni e attività, che qualificano il *creator* e vanno collocate nel tempo oltre che nello spazio. Questa area è importante perché ci permette di introdurre il principio del ruolo della persona, in un certo arco cronologico, e di valorizzare quel ruolo nei termini di una relazione fra la funzione svolta e l'entità (entità dotata di una certa funzione in un certo periodo). Nello Schema EAC-CPF l'elemento <function> è distinto da <occupation>, e la nozione di <functionRelation> è ancora differente (oltre che tipizzata attraverso l'attributo @FunctionRelationType). A sottolineare la separazione fra il concetto di funzione svolta dall'entità e la sua occupazione.[189] Nell'ontologia, mentre "occupation" è una *data property*, e quindi una stringa, "function" e "functionRelation" sono delle classi.

Al di là delle ricadute computazionali, diremo che entità CPF e identificazione univoca, spazio, tempo, relazioni tipizzate e ruoli ci introducono all'importante principio della classificazione dell'entità secondo un approccio che potremo chiamare contesto-centrico, e che ci permette di introdurre la nozione di evento. Potremmo cioè individuare nel concetto di evento quella rete di relazioni che legano la persona, dotata di un certo ruolo, a un luogo, e il luogo a un arco temporale, al fine di ricostruire un'azione.

Supponiamo di utilizzare questi parametri per astrarre il principio della descrizione di un'entità CPF. E usare gli elementi dello standard per formalizzare relazioni complesse fra agenti, cioè persone che svolgono delle azioni, e gli eventi in cui questi agenti sono coinvolti quando ricoprono un certo ruolo. E facendo qualcosa

di più, ovvero estraendo queste informazioni direttamente dal full-text di documenti testuali, che specificano il contesto dell'azione (cioè persona-ruolo-evento come entità complessa attestata in una fonte). Questo approccio si inserisce in un discorso più ampio di approfondimento delle relazioni che possono intercorrere fra agenti e documenti in cui questi agenti sono menzionati.[190]

Potremmo sintetizzare questo ragionamento nel seguente modo:

- Entità CPF. Ogni entità *creator* ha un ruolo determinato dal contesto spazio-temporale;
- Date. Le date danno consistenza alle azioni che coinvolgono le entità *creator*;
- Luoghi. I luoghi sono necessari per identificare gli spazi in cui gli eventi accadono;
- Le relazioni fra le classi appena descritte stabiliscono un approccio orientato agli eventi: le persone, dotate di un ruolo, agiscono in una certa data e in un certo luogo, stabilendo un evento. L'evento contribuisce a determinare il contesto e quindi il contesto specifica la connessione fra le classi;
- Ogni azione che viene compiuta al fine di riconoscere persone, luoghi, date ed eventi in un dato contesto, è un'attività interpretativa compiuta da un agente (colui o colei che estrae questa informazione dal full-text) sulla base della propria esperienza, o background, o sulla base dell'analisi di fonti esterne.

Per dare concretezza a questi ragionamenti sarà utile capire come modellizzare ontologicamente questa prospettiva critica. Perché prima di descrivere i dati secondo una certa modalità, che recuperi comunque la tradizione di DTDs e Schemi in uso in archivistica, è necessario stabilire un modello di osservazione del dominio che venga espresso in modo formale, ovvero ontologico. Questa riflessione sull'osservazione del *creator* in contesto, ci permette quindi di introdurre un altro lavoro di ricerca, che riguarda la creazione di un modello concettuale, denominato PRoles,[191] come ontologia che mira a ragionare sui ruoli, acquisendo informazioni dal full-text di documenti, attraverso un approccio contesto-centrico. PRoles, in particolare, importa ed estende due modelli esistenti, che abbiamo già avuto modo di menzionare: l'ontologia PRO e l'*ontology pattern* N-ary Participation.

PRO è stata concepita nell'ambito bibliografico per descrivere ruoli coinvolti nel processo di pubblicazione di un documento, benché sia facilmente estendibile a prescindere dal dominio di riferimento. In particolare, PRO definisce una classe per specificare congiuntamente l'intervallo temporale e il contesto in cui viene assunto un dato ruolo. Unitamente a un primo livello di formalizzazione delle relazioni (descrivere un agente con un certo ruolo in un certo contesto), la descrizione delle informazioni che possono essere estratte dal full-text spesso richiede un ulteriore livello di contestualizzazione, come la descrizione degli agenti che partecipano ad eventi, collocati nello spazio e nel tempo, ricoprendo un certo ruolo. Per consentire questo ulteriore livello di descrizione, PRoles include anche N-ary Participation, che permette invece di modellare ogni oggetto nei termini della sua partecipazione a un evento.

PRoles ci interessa particolarmente perché apre il ragionamento verso l'estrazione di conoscenza secondo un approccio tipicamente di derivazione filologica, ovvero l'analisi critica del testo pieno. E amplia i modelli di descrizione archivistica, e in particolare del soggetto produttore, proponendo ontologie utili tanto per arricchire record d'autorità archivistici, quanto per estendere potenzialmente l'ontologia EAC-CPF con una nuova prospettiva critica.

Per concludere. L'elemento "storia" dello standard ISAAR(CPF), campo opzionale dell'area della descrizione (5.2.2), per stessa dichiarazione dello standard, può essere riportato, nella compilazione dello strumento di corredo, in "forma narrativa".[192] Questo significa che molte informazioni, che tipicamente in un record di autorità archivistico troviamo nella parte dedicata alla storia del CPF, sono dati non strutturati, sui quali ulteriori ragionamenti, come si è cercato di avviare, potrebbero essere fatti in termini di entità persone correlate, ma anche luoghi, date ed eventi in prospettiva contesto-centrica. Come vedremo oltre (cap. 5), l'impiego dei LOD permette di rendere più granulare la descrizione archivistica,[193] agevolando la creazione di una nuova tipologia di strumento di corredo, capace di rispondere a bisogni conoscitivi che rimangono altrimenti latenti, fino almeno alla loro rielaborazione in termini *entity-centric*.

3.3 *Provenance*. Documentare i processi

Un altro concetto, tanto fondamentale quanto dibattuto in archivistica, è quello della provenienza. Anche se relativa agli oggetti digitali e finalizzata a garantire il tema dell'autenticità nel passaggio dall'analogico al digitale, questa definizione è il nostro punto di partenza:

> To validate authenticity of a preserved data object provenance is needed, i.e., the *documented history of creation, ownership, accesses, and changes* that have occurred over time for a given data object. Due to the common ancestry relations of provenance data, these data naturally form a directed graph.[194]

Primo punto di interesse è che *provenance* significa storia. Storia della creazione, della proprietà, degli accessi e delle modifiche subite da un oggetto, inteso come una qualunque fonte di informazione. Documentare la provenienza, e raccontare quindi la storia del complesso archivistico, è un elemento fondamentale di contesto, perché la provenienza è un evento e in quanto tale diversi elementi devono cooperare a documentarla: persone, luoghi, date, concetti. Riprendendo la definizione di contesto riportata in EAC-CPF, diremo che:

> Archival context information consists of information describing the circumstances under which archival materials have been created, maintained and used. This context includes, but is not limited to, the identification and characteristics of corporate bodies, persons, and families (CPF entities) who have been the creators, users, or subjects of records, as well as the relationships amongst them.[195]

Il contesto in archivistica è dunque quella serie di informazioni utili a descrivere le circostanze che hanno determinato la creazione, la gestione e l'uso del complesso documentario. Se il soggetto produttore è un elemento necessario a determinare il contesto, esso non è l'unico. Sicuramente la provenienza è un altro dato fondamentale, perché si tratta, in senso ampio, di uno degli elementi necessari a determinare il significato del complesso documentario. E dire provenienza significa, canonicamente, attribuire una qualche forma di responsabilità sui documenti, ovvero legare il soggetto produttore al complesso documentario.

Estendendo la nozione del principio di provenienza oltre l'ambito archivistico, ma acquisendo da esso il primato dell'azione, diremo che il principio di provenienza governa le pratiche della descrizione delle fonti nei termini della documentazione necessaria a garantire l'affidabilità dell'informazione. Ecco allora che, come abbiamo già

avuto modo di dire (par. 2.3), la descrizione di una risorsa culturale non può esimersi dal dichiarare la *provenance* di quella descrizione. Ovvero chi l'ha detto? Quando l'ha detto? Sulla base di quali fonti è stato asserito qualcosa? Con quale grado di certezza si afferma qualcosa?

Nell'ottica del Web semantico questa documentazione è necessaria a garantire l'autorevolezza dell'informazione, tanto per consentire inferenza consistente sui dati quanto per l'affidabilità, o anche la *trust* degli asserti. Problema che, in particolare nel dominio dei LOD, fa la differenza: un asserto privo di contesto non necessariamente è sbagliato, ma sicuramente è, filologicamente, meno affidabile.

Riprendiamo il modello HiCo, che a sua volta nasce dalla riflessione condotta con l'elaborazione di Proles. E proviamo a riflettere sul principio dell'interpretazione come atto condotto su un frammento del testo di un'edizione, recuperando il già citato esempio delle lettere di Vespasiano da Bisticci. Nel caso di un'affermazione, o anche un'interpretazione, dell'editore critico, è stato possibile formalizzare un criterio, facilmente condivisibile dalla comunità di riferimento, per stabilire quando un asserto sia considerato affidabile, identificando tre situazioni principali: la citazione della letteratura sull'argomento, la consultazione di una fonte autorevole, una lettura personale dell'interprete per avvalorare la tesi proposta. In questo caso, l'attribuzione di un ruolo a una persona citata nel testo pieno di una lettera della raccolta epistolare è formalizzabile come un atto interpretativo, effettuato dall'editore critico delle lettere e trasformato in *statement* RDF da un altro agente, un atto che viene tipizzato ed è caratterizzato da informazioni aggiuntive che l'editore ha fornito per validare l'asserzione. Nell'esempio riportato di seguito, l'interpretazione è avvalorata dalla consultazione di un manoscritto autografo di una lettera, corroborata da una lettura personale dell'editore e sostenuta da altri editori della letteratura sull'argomento (Fig. 6).

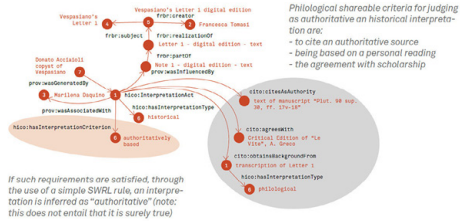
Philological shareable criteria for judging as authoritative an historical interpretation are:
- to cite an authoritative source
- being based on a personal reading
- the agreement with scholarship
Vespasiano's Letter 1
Vespasiano's Letter 1 digital edition
frbr:creator
Francesco Tomasi
frbr:subject
frbr:realizationOf
Letter 1 - digital edition - text
frbr:partOf
Note 1 - digital edition - text
prov:wasInfluencedBy
Donato Acciaioli copyist of Vespasiano
prov:wasGeneratedBy
Marilena Daquino
hico:InterpretationAct
hico:hasInterpretationType
historical
prov:wasAssociatedWith
hico:hasInterpretationCriterion
authoritatively based
cito:citesAsAuthority
text of manuscript "Plut. 90 sup. 30, ff. 17v-18"
cito:agreesWith
Critical Edition of "Le Vite", A. Greco
cito:obtainsBackgroundFrom
transcription of Letter 1
hico:hasInterpretationType
philological
If such requirements are satisfied, through the use of a simple SWRL rule, an interpretation is inferred as "authoritative" (note: this does not entail that it is surely true)

Figura 6 - Nella prima nota dell'edizione digitale della prima lettera (5) della raccolta, Francesca Tomasi (l'editore critico, 2) afferma che Donato Acciaiuoli ha ricoperto il ruolo di copista per Vespasiano (7). Questa affermazione (1) è corroborata dalla lettura personale della studiosa e da un'altra edizione critica (ed. Greco). L'editore critico ha basato la sua interpretazione su un manoscritto conservato presso la BML di Firenze (Plut. 90 sup. 30), lettera dell'Acciaiuoli a Vespasiano. Nella nota si rimanda al ms. Magliabechiano VIII 1390, autografo dell'Acciaiuoli. Chi crea lo *statement*, Marilena Daquino, può quindi considerare questa interpretazione come autorevole, perché documentata e arricchita attraverso i necessari elementi di contesto

Tramite questi tre principi basilari per la definizione di un'interpretazione storica autorevole – cioè non necessariamente vera, ma solidamente documentata – è possibile inferire una nuova proprietà con un dato valore (*authoritatively-based*) per caratterizzare ulteriormente l'entità rappresentante l'interpretazione.

Il punto è che ogni azione va documentata o anche ogni interpretazione condotta sui dati va descritta. Con la consapevolezza che gli asserti variano al variare del contesto. E allora, come abbiamo già detto, non potremo avere un unico *knowledge graph*, ma un *knowledge graph* che varia al variare del punto di vista dell'osservatore o anche a seconda della provenienza dell'asserto, con l'idea di far convivere posizioni diverse, e magari contrastanti, sullo stesso oggetto osservato.

Il problema principale di documentare situazioni come quelle descritte è che RDF si basa sulla nozione di asserto nella forma di una tripla. È evidente dunque che è necessario mettere in dialogo fra di loro triple relative alla stessa entità concettuale, in particolare quando gli asserti siano appunto complessi, ovvero richiedano l'aggiunta, all'asserto primario, di informazioni sulla provenienza e, in generale, sui contesti.

Una risposta ci arriva dal Web, ed è prima concettuale e poi

tecnica: i *named graphs*[196] o grafi nominati (ovvero identificati univocamente). Si tratta di un concetto chiave dell'architettura del Web semantico: un insieme di dichiarazioni RDF (*statement* primario più *statements* contestuali) può essere identificato utilizzando un URI globale dell'insieme, rendendo così possibili descrizioni del contesto, della provenienza e altre informazioni o metadati. Si garantisce, o meglio si forniscono gli elementi sufficienti per garantire, di conseguenza l'affidabilità degli asserti.[197]

Per riflettere sull'uso dei *named graphs* in un caso in cui sia necessario documentare la *provenance*, torniamo a un esempio già introdotto (si veda par. 2.4): il lavoro condotto sull'edizione digitale semantica del quaderno di Paolo Bufalini (1981-1991).

Paolo Bufalini è stato membro del PCI e Senatore della Repubblica dal 1963 al 1992. Latinista e raffinato traduttore di Orazio, fra il 1981 e il 1991 Bufalini ha tenuto un quaderno di appunti privati composto da 145 pagine rilegate e 2 carte sciolte. Il quaderno contiene frammenti testuali relativi a momenti significativi della vita intellettuale e sociale dell'autore: citazioni tratte da testi letterari, note personali e commenti, narrazioni di eventi, prove di traduzione dal latino all'italiano. In particolare, data la natura privata del quaderno, le relazioni fra i frammenti testuali (testo-testo, testo-nota, testo-traduzione e traduzione-traduzione) non sono rese esplicite da Bufalini, così come non sempre viene dichiarata la fonte dei testi citati. Alla morte di Paolo Bufalini, il quaderno è stato donato dagli eredi al Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna, che si è occupato di trascrivere gli appunti e di identificare le fonti delle 196 citazioni presenti. I testi, accompagnati da traduzioni e annotazioni, provengono dai classici della letteratura italiana e latina (Dante, Manzoni, Petrarca, Carducci, Virgilio, Orazio, Tacito, Cicerone, Ennio, Marziale ecc.) e delle letterature europee (Flaubert, Hegel, Shakespeare, Mann, Yourcenar ecc.).

L'obiettivo che il gruppo di ricerca si è posto, è stato quello di procedere alla realizzazione di un'edizione digitale semantica. Scopo dell'edizione è stato la valorizzazione degli aspetti intra e intertestuali veicolati dalla fitta rete di citazioni, traduzioni e commenti di mano dell'autore.

L'edizione del quaderno è caratterizzata da due livelli attraverso cui tradurre il principio dell'interpretazione: c'è l'interpretazione condotta dallo stesso autore (Bufalini), il quale annota testi e

sottolinea influenze fra opere e/o autori, stabilisce potenziali collegamenti e interconnessioni, e c'è l'interpretazione degli editori, che individuano e analizzano la trama di tali relazioni, anche quando non dichiarate in modo esplicito dall'autore, formalizzandole.

Al fine di distinguere i diversi piani e associare ciascuna asserzione RDF (quella dell'autore e quella dell'editore) alla rispettiva provenienza, si è scelto di utilizzare i grafi nominati e, in particolare:

- le asserzioni che riguardano il testo, quali commenti, persone e testi citati, riferimenti bibliografici, fanno parte di un grafo (grafo del testo);
- le asserzioni RDF che riguardano le relazioni fra persone e fra opere citate costituiscono grafi separati (grafo delle interpretazioni).

A questo modello è stato aggiunto un ulteriore livello, necessario a fornire informazioni circa l'operazione di pubblicazione dei *named graphs*. Si tratta delle *nanopublications*,[198] un modello che consente di rappresentare le asserzioni, il contesto di riferimento e i dati di pubblicazione in un unico grafo.

Secondo il modello delle *nanopublication* è necessario provvedere a tre elementi base: 1. l'asserzione (*assertion*) che stabilisce lo *statement* base; 2. la *provenance* che aggiunge i metadati, attribuendo un primo stato di contesto all'asserto, consentendo di dichiarare anche i metodi utilizzati per generare l'asserto e il suo grado di autorevolezza; 3. le informazioni di pubblicazione (*publication information*) che riguardano i metadati della *nanopublication* come un unico *statement*, e documentano tanto l'asserzione che la provenienza, aggiungendo informazioni relative alla pubblicazione del dataset.

Ogni grafo rappresentante gli asserti (ad es. la relazione tra due autori) è collegato quindi a due ulteriori grafi: il primo contenente la provenienza delle asserzioni estratte dal quaderno (ad es. Bufalini cita il testo di un autore terzo che afferma esistere una relazione tra due autori); il secondo contenente le informazioni inerenti alla pubblicazione dell'edizione digitale (il Dipartimento di Filologia Classica e Italianistica). Un quarto grafo identifica la *nanopublication* e la collega ai precedenti grafi.

Questa strategia permette di citare ogni singolo asserto (ossia ogni interpretazione effettuata sul testo) quale entità autonoma dall'edizione in cui nasce, ma sempre riconducibile a essa mediante la sua *provenance*. In ultimo, una *integrity key*[199] associata alle *nanopublication* consente di fare riferimento a una versione stabile dell'edizione, assicurandone la citabilità a lungo termine. Tema fondamentale quello della citabilità, necessario per far fronte a diversi aspetti che ci interessano: 1) il problema della scientificità delle pubblicazioni digitali; 2) l'annosa problematica della preservazione concettuale dei dati; 3) il tema, altrettanto importante, dell'autorevolezza dei contenuti digitali come fonti di informazione di qualità.

3.4 Dall'albero al grafo. Il ruolo di RiC

È naturale pensare al contesto in termini di posizionamento. Posizionamento di una parola nel testo, di un libro su uno scaffale, di un quadro in un museo, di un documento in un complesso documentario. La posizione stabilisce il ruolo dell'oggetto nel contesto, determinato dalla relazione di quell'oggetto con gli oggetti che lo circondano. E in campo archivistico la posizione del documento è un elemento fondamentale della descrizione.

La descrizione archivistica è tradizionalmente di tipo gerarchico. È quindi l'albero la struttura dati che governa da sempre la modalità di organizzazione fisica delle carte, e che stabilisce le regole cui attenersi per la creazione di strumenti di corredo archivistici.

Lo standard ISAD ci dice chiaramente che la descrizione è relativa al livello. E l'importanza della gerarchia in archivistica è determinata dal principio che vede nella creazione di mezzi di corredo, in particolare primari, l'esigenza del rispetto delle relazioni di subordinazione/sovra-ordinazione fra i livelli descrittivi del complesso documentario (tradizionalmente classificabili in: fondo, subfondo; serie, subserie; fascicolo o unità archivistica; unità documentaria).

La gerarchia permette di ragionare in forma tipicamente tassonomica e quindi di percorrere l'albero. Le possibilità di interrogare questa struttura dati è legata al rapporto padre/figlio degli elementi della gerarchia. Quello gerarchico è un modello

storicamente significativo; non è un caso che in computer science il modello di database gerarchico sia stato il primo ad affermarsi sul mercato. Un altro esempio di modello gerarchico è quello del file system, in uso sulla totalità dei computer. E tutti siamo abituati a navigare il disco rigido in forma gerarchica secondo una struttura per cartelle, sottocartelle e file. Questo significa che una struttura dati ad albero prevede un approccio strettamente verticale. Con tutte le limitazioni che ne derivano.

Ma, come provocatoriamente afferma Michetti, "è poi tanto pacifico che l'albero rispecchi l'archivio?".[200] Perché in fondo la descrizione archivistica richiede una serie di interconnessioni orizzontali e trasversali che servono ad arricchire l'esperienza conoscitiva di chi consulta il complesso documentario. E le tecnologie servono proprio per superare la fissità delle raccolte analogiche organizzate per faldoni, fascicoli, registri ecc. per abbracciare le relazioni che esistono fra le entità coinvolte nell'insieme dei documenti. Affinché un elemento di descrizione restituisca conoscenza dobbiamo necessariamente andare oltre la gerarchia, ovvero abbracciare la rete dei collegamenti trasversali. Il contesto è tanto verticale (l'albero inteso in termini tassonomici), quanto orizzontale, (le relazioni di pari livello). Ed è in una riflessione simile che si colloca il passaggio dall'albero strettamente gerarchico a un modello a grafo etichettato e tipizzato, ovvero diremo anche dalla gerarchia alla multidimensionalità delle relazioni.

Che il grafo sia un modello utile per la descrizione archivistica è un fatto. Ma alla luce del Web semantico possiamo dire di più. La rete è il modello scelto dal WWW. E il grafo è la struttura dati che sta alla base dei LOD, ovvero anche potremmo dire che RDF è un formato di descrizione delle risorse, basato su un modello dei dati a grafo. E che il *knowledge graph* sia uno dei concetti più attestati nella recente letteratura sul tema Semantic Web ci fa capire che non parliamo solo di una modalità di rappresentazione dei dati, ma di un sistema di analisi della conoscenza. Ovvero, interrogare i dati all'interno di un grafo è diverso dall'interrogare una base di conoscenza in un DBMS relazionale.

Dati strutturati e non strutturati possono convivere in un grafo della conoscenza che esprime l'interpretazione condotta sui dati al fine della creazione dei sistemi di interconnessione più appropriati a descrivere, attraverso modelli ontologici, la realtà osservata. Ed è

dopo aver individuato le relazioni interne che è possibile arricchire il proprio dominio con tutti i dati che il *cloud* mette a disposizione e che sono necessari a costruire la base di conoscenza, valorizzando il proprio *knowledge graph* di progetto e facendolo dialogare con altri grafi.

Potremmo anche dire che il grafo, che dal punto di vista della computer science possiamo definire come una generalizzazione della struttura ad albero, introduce, con la nozione di numero arbitrario di nodi vicini e di cicli, il principio della relazione fra i dati e il loro contesto, agevolando la costruzione di sistemi di interrogazione che permettano l'esplorazione dei dati semantici, ovvero basati su ontologie, muovendo da un approccio basato su parole chiave (*keyword-based*) a uno basato sulla ricerca del significato (*meaning-based*).

Non è un caso se, più di recente, dai cosiddetti *triplestore* (datasets di triple) si sia passati ai *quadstore*, che recuperano il concetto dei *named graphs* descritto nel paragrafo precedente, arricchendo di contesto la tripla. Il *quadstore* aggiunge alla tripla (soggetto-predicato-oggetto) la nozione di grafo nominato come componente indispensabile alla rappresentazione del significato. Ma la complessità della computabilità dei *quadstore* è un problema e la riflessione sui *named graphs* si è ampliata. Il recente interesse della comunità scientifica per RDF* (RDF Star)[201] che consente di aggiungere metadati al grafo (fra cui, ad es., la *provenance*, ma anche aspetti temporali, spaziali e in generale di contesto), mantenendo comunque una struttura per triple, e rendendo più efficiente la reificazione con una nuova sintassi, è indice di una tendenza.[202]

E non è un caso poi se i database a grafo, una delle prime sperimentazioni nel settore,[203] siano oggi un altro strumento utilizzato per la creazione di basi di conoscenza nell'ambito dei Web semantico (e per la pubblicazione di LOD in particolare). I database a grafo[204] – alternativa tanto ai database relazionali basati su strutture a tabelle, quanto ai NoSQL database solo *document-centric*[205] –, richiamano il sistema cosiddetto orientato agli oggetti (*object-oriented*), dove i nodi non sono solo tipi di dato, ma istanze del grafo. Ogni istanza ha proprietà che la descrivono (*datatype properties*) e proprietà che collegano l'istanza ad altri oggetti (*object properties*), in modo tale che le *datatype properties* sono integrate per formare una descrizione dell'istanza, mentre le *object properties* sono

trattate come archi che connettono diverse istanze. I nodi in più non sono semplici stringhe ma oggetti, dotati di una molteplicità di *datatype properties*. Concetti di modellazione noti a chi abbia familiarità con le ontologie e con l'Ontology Web Language (OWL), eletto dal W3C come il linguaggio "designed to represent rich and complex knowledge about things, groups of things, and relations between things".[206]

Il grafo è quindi un modello di dati che arricchisce di espressività la descrizione degli oggetti del patrimonio culturale e che ha una serie di importanti implicazioni sulla capacità di acquisire nuova conoscenza dai dati stessi, così rappresentati nella forma di oggetti; oggetti che possono anche essere testi pieni (secondo il principio *text-as-a-graph*).[207]

Ci avviamo allora alla chiusura con un'ultima, importante riflessione. Che l'archivistica abbia sposato la nozione di molteplicità di contesti è indubbio, e che i contesti siano rappresentabili attraverso strutture a grafo, esprimibili come relazioni fra un soggetto e un oggetto, è altrettanto indubbio.

L'ultimissima release nel settore degli standard archivistici conferma questa tendenza. Le istanze di rinnovamento concettuale determinate dal Web semantico sono infatti le ragioni dell'avvio di un processo di revisione dei quattro standard archivistici (ISAD, ISAAR(CPF), ISDIAH e ISDF). Nel 2012 l'EGAD (Expert Group on Archival Description) è stato incaricato di progettare un nuovo modello descrittivo. Quattro anni dopo è stato elaborato il già menzionato modello concettuale Records in Contexts (RiC-CM). RiC-CM cerca di superare il tradizionale approccio archivistico gerarchico, per abbracciare la descrizione su un piano multi-dimensionale, più adeguato ai requisiti del Web semantico. RiC-CM cerca cioè di adottare un nuovo modello, cioè il grafo, per sostituirlo a un approccio che tradizionalmente usa l'albero gerarchico come struttura di riferimento. E la ragione è strettamente legata al linguaggio di rappresentazione degli strumenti archivistici. Mentre i primi tentativi in campo archivistico, a seguito degli standard metodologici, sono stati di elaborare (e quindi adottare) DTD prima e Schemi poi (cioè EAD e EAC(CPF)) per usare una sintassi, a base gerarchica ad albero, come l'XML, nel momento in cui il Web semantico ha optato per adottare il sistema RDF, che ragiona nella dimensione della relazione espressa dalla rete o anche dal network, ovvero usa il modello del

grafo, anche DTD e Schemi sono stati abbandonati in favore dell'ontologia.

> RiC Relations diverge conceptually from XML-based standards like EAD and EAC that rely on structural hierarchy to define connections between, for example, records to series to fonds. Instead, RiC Relations are based on a graph model, similar to the Resource Description Framework (RDF), which allows for simpler and more flexible connections.[208]

RiC ha massimizzato lo sforzo nella classificazione delle entità coinvolte nel processo di descrizione archivistica, che ha al suo centro la creazione del record: agenti, ma anche concetti/soggetti, luoghi e date sono i punti di accesso all'informazione, e vengono arricchiti delle nozioni di funzione, attività, mandato, occupazione e posizione, che caratterizzano l'agente e sono fondamentali contesti di creazione dei record archivistici, necessari a garantire autenticità e autorevolezza dei documenti, oltre a rafforzare il fondamentale principio di provenienza. Questi sono dunque i diversi strumenti attraverso cui arricchire la descrizione del patrimonio archivistico oltre il canonico principio della separazione della descrizione del complesso documentario dalla descrizione del soggetto produttore. Nuove entità, oltre al *creator* CPF, contribuiscono a dare quei contesti necessari a garantire espressività del dato. E l'entità, come i LOD ci insegnano, è il cuore del processo di modellazione, prima di tutto a un livello astratto e poi calando l'astrazione sulle istanze attraverso proprietà semanticamente esplicite (i predicati).

RiC-CM versione 2[209] è stata naturalmente la base di riflessione concettuale necessaria a realizzare la già menzionata ontologia di dominio RiC, ovvero RiC-O,[210] rilasciata, nella versione corrente, nel febbraio 2021.

RiC-O traduce in OWL la concettualizzazione di RiC-CM e propone una serie di sottoclassi dipendenti da una generica classe Thing (idea, concetto, oggetto o evento), per adeguare il modello agli approcci ontologici (ad es. owl:Thing, la classe radice di tutte le ontologie espresse in OWL). Record Resource e Agent sono le due classi principali, che si arricchiscono, attraverso opportuni predicati, di una serie di nozioni fondamentali alla contestualizzazione: evento, ovvero qualcosa che accade in una dimensione spazio-temporale e che richiama il concetto di attività, classificabile come un certo tipo di evento; regola, che introduce il concetto di mandato (un certo tipo di regola), ovvero si riferisce

alle condizioni per cui una certa attività o funzione, svolta o ricoperta dagli agenti, è avvenuta in un certo specifico contesto; la data, per la componente contestuale di tipo temporale; luogo, per la componente contestuale di tipo spaziale o geografico; instanziazione, ovvero l'oggetto fisico, l'item diremo, o un suo supporto digitale in un qualche formato (il cambio di formato determina una nuova instanziazione del record). E la *provenance* gestisce il rapporto della Record Resource, e della sua Instantiation, con l'Agent.[211]

Nonostante le numerose critiche e le importanti discussioni sul modello,[212] RiC incarna la nuova tendenza nella rappresentazione della conoscenza: assumere la teoria alla base del Web semantico per ripensare ai criteri di descrizione dei dati attraverso il principio della costruzione di relazioni trasversali (il grafo) fra unità di contenuto (le entità del grafo). Potremmo dire che RiC vuole affermare l'istanza di rinnovamento della descrizione archivistica verso l'interoperabilità semantica", dove:

> The concept of description level is not envisaged in this new model where hierarchy is not a determining factor; instead it is the unique representations based on exact **relations** between independent entities that are able to capture **dissimilar and complex creation contexts**.[213]

Un'interessante proposta del modello è ovviamente il ragionamento sulle relazioni, e in particolare sul problema dell'arricchimento della tripla con nuovi attributi. Al momento RiC-O propone un sistema di classi per implementare la sezione Relazioni di RiC-CM. Sebbene, dicano le specifiche, anche queste relazioni siano rappresentate come semplici (e binarie) proprietà dell'oggetto (ad es. hasProvenance che corrisponde alla relazione RiC-R026), potrebbe essere necessario assegnare attributi diversi a una relazione, ad es. una data, un grado di certezza o una descrizione, ma anche il riferimento a una fonte, come è già possibile in un file XML/EAC-CPF. Uno dei metodi standard disponibili, continuano le specifiche, per rappresentare una relazione così documentata in RDF per ora è usare una classe. Le specifiche RDF* e SPARQL*, sviluppate dal W3C RDF-DEV Community Group,[214] forniscono un metodo molto più semplice (che consente di considerare una tripla come soggetto o oggetto di un'altra tripla), già utilizzato, ma ancora non standard W3C. Così, ad es., in RiC-O esiste una classe AgentOriginationRelation. Questa

classe può connettere da uno a molti agenti a da uno a molti record risorse o istanze create o accumulate e ha alcune proprietà specifiche dell'oggetto (certezza, data, descrizione, origine/fonte).[215]

Peccato solo che RiC-O rifletta di nuovo una tendenza, che si auspica voler superare nella modellazione della conoscenza: il mancato sforzo di allineamento con i modelli esistenti, non tanto, o meglio non solo, sul piano terminologico, che pur giova, quanto su quello concettuale. Un vero approccio evento centrico, come può essere quello perseguito da una certa modalità di rappresentazione della conoscenza in EDM, o sicuramente quello che identifica e qualifica il cuore del modello CIDOC CRM, è necessario per dare una giusta descrizione dei dati, che altrimenti rimangono privi di quei necessari contesti utili a dare espressività al dato stesso.[216]

Solo quando l'allineamento concettuale, prima che terminologico, dei nuovi vocabolari del Web semantico sarà un'attività condivisa e collaborata, allora il grafo potrà davvero esaudire il suo potere di nuovo strumento di organizzazione della conoscenza.

Vediamo allora come il mondo delle biblioteche ha risposto a questa esigenza, ripartendo con FRBR nel prossimo capitolo.

179 Michael Gavin, *Vector Semantics, William Empson, and the Study of Ambiguity*, "Critical Inquiry", 44 (2018), 4, p. 641-673, https://web.stanford.edu/~jurafsky/slp3/6.pdf.

180 Ci si riferisce qui ai quattro livelli della piramide DIKW (Data, Information, Knowledge, Wisdom) per cui si veda Russell Ackoff, *From Data to Wisdom*, "Journal of Applied Systems Analysis", 16 (1989), p. 3-9.

181 Il riferimento è naturalmente a Roman Jakobson con il suo *Linguistics and Poetics*, per cui si veda *Style in language*, edited by Thomas A. Sebeok, New York, John Wiley & Sons, 1960, p. 350-377.

182 Shelley Sweeney, *Provenance of Archival Materials*, in *Encyclopedia of Library and Information Sciences*, Third Edition, edited by Marcia J. Bates, Mary Niles Maack, vol. VI, p. 4315-4323, Boca Raton Florida, CRC Press, 2010.

183 Questi sono i principi che stanno a fondamento anche, ovviamente, della costruzione di strumenti di corredo archivistici, per cui si veda ICA/CDS, *Linee-guida per l'elaborazione e la presentazione di strumenti di ricerca*,

trad. it. 2004, a cura di Francesca Ricci, http://media.regesta.com/dm_0/ANAI/anaiCMS/ANAI/000/0111/ANAI.000.0111.0007.pdf.

184 Si veda in particolare *Authority control: definizione ed esperienze internazionali: atti del convegno internazionale* (Firenze, 10-12 febbraio 2003), a cura di Mauro Guerrini, Barbara B. Tillett, Firenze, Firenze University Press; Roma, AIB 2003.

185 Definizione canonica, che si può leggere in Direzione Generale Archivi (DGA), *Glossario*, http://www.archivi.beniculturali.it/index.php/abc-degli-archivi/glossario.

186 Si vada in particolare Daniel Pitti, *Creator Description: Encoded Archival Context*, in *Authority control in organizing and accessing information: definition and international experience*, edited by Arlene G. Taylor, et al., Binghamton N.Y., Haworth Information Press, 2004, p. 201-226.

187 Una proposta in Valentina Anita Carriero, Marilena Daquino, Francesca Tomasi, *Convergenze semantiche tra musei, archivi e biblioteche. Ontologie per le relazioni interpersonali,* "JLIS.it", 10 (2019), 1, p. 72-91.

188 Per cui si veda Silvia Mazzini, Francesca Ricci, *EAC-CPF Ontology and Linked Archival Data*, in *Proceedings of the 1st International Workshop on Semantic Digital Archives (SDA 2011)*, "Ceur Workshop Proceedings", 801 (2011), p. 72-81, http://ceur-ws.org/Vol-801/paper6.pdf.

189 Non a caso sul concetto di funzione l'archivistica ha ragionato, tanto da creare, nel 2007, uno standard ad hoc che è il già citato ISDF.

190 Particolarmente interessante per questa riflessione il modello "factoid", in cui è possibile fare asserti su persone, così come estratti da una fonte di riferimento: "a factoid is a kind of prosopographical assertion that centers on statements made by an historical source. It is a structured interpretation of something that an historical source says about an individual". Si veda John Bradley, Michele Pasin, *Factoid-based prosopography and computer ontologies: towards an integrated approach*, "Digital Scholarship in the Humanities", 30 (2015), 1, p. 86-97, http://doi.org/10.1093/llc/fqt037.

191 Marilena Daquino, Silvio Peroni, Francesca Tomasi, Fabio Vitali, *Political Roles Ontology (PRoles): enhancing archival authority records through Semantic Web technologies*, "Procedia Computer Science", 38 (2014), Elsevier, p. 60-67.

192 ISAAR-CPF, cit., p. 231.

193 Jinfang Niu, *Linked Data for Archives*, "Archivaria", 82 (2016), p. 83-110, https://archivaria.ca/index.php/archivaria/article/view/13582.

194 Michael Factor, et al., *Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware Storage*, "First workshop on the Theory and Practice of Provenance", 2009, 6, p. 1-10, https://dl.acm.org/doi/10.5555/1525932.1525938.

195 EAC(CPF), *Tag Library*, https://eac.staatsbibliothek-berlin.de/schema/taglibrary/cpfTagLibrary2019_EN.html.

196 Jeremy J. Carroll, Christian Bizer, Pat Hayes, Patrick Stickler, *Named*

graphs, "Journal of Web Semantics", 3 (2005), 4, p. 247-267.

197 Jeremy J. Carroll, Christian Bizer, Pat Hayes, Patrick Stickler, *Named graphs, provenance and trust*, in *Proceedings of the 14th international conference on World Wide Web (WWW '05)*, NY, ACM, 2005, p. 613-622, https://doi.org/10.1145/1060745.1060835.

198 Le *Guidelines* si possono leggere alla URL: http://nanopub.org/guidelines. Si veda in particolare Paul Groth, Andrew Gibson, Jan Velterop, *The Anatomy of a Nanopublication*, "Information Services & Use", 30 (2010), 1-2, p. 51-56, https://doi.org/10.3233/ISU-2010-0613.

199 *Nanopublication Guidelines*, http://www.nanopub.org/2013/WD-guidelines-20131215/#integrity-key.

200 Giovanni Michetti, *Ma è poi tanto pacifico che l'archivio rispecchi l'albero?*, "Archivi e Computer", 14 (2009), 1, p. 85-95. Dello stesso autore *Archives are not a Tree*, in *The Memory of the World in the Digital Age: Digitization and Preservation*, edited by Luciana Duranti, Elizabeth Shaffer, Vancouver, UNESCO, 2013, p. 1002-1010.

201 Olaf Hartig, *Foundations of RDF* and SPARQL*: an alternative approach to statement-level metadata in RDF*, in *11th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)*, "CEUR Workshop Proceedings", 1912 (2017), http://ceur-ws.org/Vol-1912/paper12.pdf.

202 Si veda il recente *survey* delle diverse soluzioni per gestire le triple arricchite di contesto (oltre il canonico approccio dei *named graphs*): Leslie F. Sikos, Dean Philp, *Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs*, "Data Science and Engineering", 5 (2020), p. 293-316.

203 Sarà interessante ricordare a questo proposito il progetto CLIO/kλειωm creato da Manfred Thaller, *KLEIO, A Data Base System for Historical Research*, version 1.1.1/MS DOS, Gôttingen, Max -Planck Institut, 1987. Si veda, in particolare, Manfred Thaller, *Clio – ein datenbankorientiertes system für die historischen Wissenschaften: Fortschreibungsbericht*, "Historical Social Research", 12 (1987), 1, p. 88-91.

204 Fra i più noti: Neo4j, Amazon Neptune e Blazegraph (quest'ultimo è sia un triplestore che un database a grafo).

205 Un esempio è MongoDB, https://www.mongodb.com.

206 Ontology Web Language (OWL), https://www.w3.org/OWL.

207 Per questa discussione si veda in particolare il progetto *Codex*, realizzato con Neo4j e basato su un sistema di annotazione *stand-off* per la gestione delle proprietà. Lian Neill, *Building a Graph of History with The Codex*, 2020, https://neo4j.com/blog/building-graph-history-codex. Di interesse il tool SPEEDy per la *stand-off annotation* e per cui si veda Lian Neill, Desmond Schmidt, *SPEEDy. A Practical Editor for Texts Annotated with Standoff Properties*, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, cit., p. 45-54.

208 RiC-CM, *Consultation Draft 2019*, p. 51, https://www.ica.org/sites/

default/files/ric-cm-0.2_preview.pdf.

209 Nella versione RiC-CM 0.2 del 2021 le entità core sono state ripensate nei termini di Record Resource, Instantiation, Agent, Event, Rule, Date e Place. E sono le classi principali dell'ontologia.

210 È disponibile un *repository* che raccoglie, oltre all'ontologia in OWL, la documentazione e gli esempi https://github.com/ICA-EGAD/RiC-O.

211 Nella pagina di documentazione dell'ontologia si veda in particolare l'immagine che rappresenta il "diagram shows the main RiC-CM v0.2 entities and a few relations between them", https://raw.githubusercontent.com/ICA-EGAD/RiC-O/master/diagrams/diagrams_v0-2/RiC-CM-overview/diagram_RiC-CM-overview-RiC-v0-2.jpg.

212 Si veda, ad es., la "call for comments" lanciata in occasione della presentazione del modello: Daniel Pitti, Bill Stockting, Florence Clavaud, *Presentation of the standard "Records in Context"*, Paris, International Council on Archives, 2016, https://www.ica.org/en/presentation-standard-records-context-thursday-8-september-1500. Ma si veda anche *Records in Contexts. A conceptual model for archival description. Il contributo italiano*, "Il mondo degli archivi", Quaderni (2017), 2, http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n2.pdf, e ancora Pierluigi Feliciati, *Archives in a Graph. The Records in Contexts Ontology within the framework of standards and practices of Archival Description*, "JLIS", 12 (2021), 1, p. 92-101, http://dx.doi.org/10.4403/jlis.it-12675.

213 Dunia Llanes-Padrón, Juan-Antonio Pastor-Sánchez, *Records in contexts: the road of archives to semantic interoperability*, "Program", 51 (2017), 4, p. 387-405, https://doi.org/10.1108/PROG-03-2017-0021. Grassetto nostro.

214 Si veda in particolare *RDF-star and SPARQL-star*. Draft Community Group Report, 01 July 2021, https://w3c.github.io/rdf-star/cg-spec/2021-07-01.html.

215 Tradotto con modifiche dalla sezione *From RiC-CM to RiC-O*, https://www.ica.org/standards/RiC/RiC-O_v0-2.html.

216 Va detto che il mapping (terminologico) è comunque fra le attività future del gruppo di lavoro: "adding suggestions of mappings (in rico:closeTo) and OWL equivalences between some classes or properties and components in other models (among which - this is not an exhaustive list- CIDOC CRM, IFLA LRM, PREMIS, PROV-O, Wikidata and Schema.org)", https://www.ica.org/standards/RiC/RiC-O_v0-2.html.

4. RAGIONARE A LIVELLI OLTRE L'UNIVERSO BIBLIOGRAFICO

4.1 Livelli ed entità

Livelli ed entità rappresentano due concetti importanti nel Web semantico, e in particolare nella prospettiva della creazione di dati aperti e interconnessi ad alto tasso di espressività.

L'entità è il *core* dei LOD perché rappresenta ogni cosa che può essere soggetto (o oggetto) di una tripla. I livelli sono determinanti nel Web semantico (o meglio diremo nel suo *stack*) per il principio della stratificazione dell'approccio ai dati – dalla standardizzazione della codifica delle stringhe di caratteri al formalismo per esprimerle, salendo verso la sintassi di rappresentazione, per arrivare alla semantica e infine all'inferenza come procedimento logico sui dati.

Livelli ed entità richiamano naturalmente il modello FRBR che, nato in contesto bibliografico, è diventato, come vedremo oltre, un *framework* teorico trasversale rispetto al dominio di origine.

Secondo FRBR il modello WEMI (Work, Expression, Manifestation, Item) permette di ragionare sugli oggetti culturali adottando quattro lenti interpretative di osservazione, ciascuna delle quali attribuisce un significato potenzialmente diverso al valore dello stesso descrittore (ad es. ci sarà una data, ma anche un luogo o un autore, dell'opera, una data dell'espressione, una data della manifestazione e una data dello specifico oggetto fisico). Le entità di Gruppo 1 determinano quindi i livelli di osservazione, e gli attributi per ogni entità stabiliscono i metadati descrittivi, mentre le relazioni determinano i collegamenti fra le entità. Ovvero in FRBR le entità W, E, M, I sono i livelli. Ma ci sono anche le entità di Gruppo 2,[217] ovvero gli agenti (persone o enti) che hanno una

qualche relazione (responsabilità intellettuale, produzione, diffusione, tutela) con i livelli di Gruppo 1. Le entità di Gruppo 3218 sono i soggetti delle entità di Gruppo 1: concetti, oggetti, eventi e luoghi (Fig. 7). Tutto quello che serve per una descrizione espressiva degli oggetti culturali: contenuto, e non solo contenitore, paternità, ovvero contesto, e soggetto, ovviamente fra di loro in relazione.

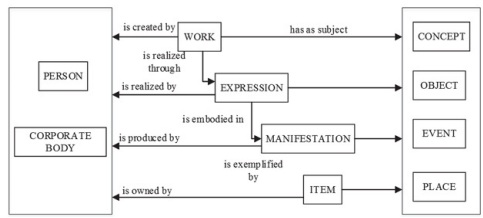

Figura 7 - Il modello FRBR. Le entità dei diversi Gruppi e le relazioni di Gruppo 2 e Gruppo 3 con i diversi livelli delle entità di Gruppo 1.
Tratto da https://commons.wikimedia.org/wiki/File:FRBR-Group-2-entities-and-relations.svg

Diremo che la novità rilevante del modello concettuale introdotto da FRBR, rispetto alla tradizione delle pratiche catalografiche tradizionali, sta nell'aver spostato il punto di attenzione del processo di catalogazione dai criteri di compilazione della scheda catalografica all'identificazione della natura del contenuto informativo presente in una scheda catalografica, distinguendo chiaramente la descrizione del contenuto intellettuale dalla descrizione dell'oggetto materiale.

In un catalogo che usi FRBR come strategia di rappresentazione degli oggetti, le diverse tipologie di dato possono essere organizzate durante la catalogazione in accordo con la struttura prevista dal modello descrittivo e potrebbero così confluire nei metadati specifici di ciascuna entità del dominio. È così che la stringa "Roma" può indicare parte del titolo di un'opera, il luogo di edizione di un libro o la città natale di un autore.

In altre parole, diremo che il modello concettuale FRBR si pone l'obiettivo di strutturare i contenuti dell'informazione bibliografica in modo da rispondere efficacemente alle esigenze di consultazione degli utenti finali. L'articolazione gerarchica del modello FRBR, procedendo dal piano astratto dell'opera fino al piano materiale dell'esemplare, offre la possibilità di scegliere il punto di ingresso

da cui avviare l'esplorazione dei contenuti. Attraverso i rinvii tra le entità è possibile muoversi tra i diversi livelli: il riferimento bibliografico di un'opera permette di spostarsi dal contenuto (l'opera), al contenitore (il volume che la contiene).

È un processo di natura epistemologia, ovvero è l'esigenza di tradurre l'interpretazione degli oggetti culturali attraverso un filtro di osservazione sfaccettato sulle diverse entità coinvolte in un oggetto culturale e i diversi livelli attraverso cui ogni oggetto dichiara i suoi elementi descrittivi.

Per la nostra argomentazione dunque le entità di Gruppo 1 del modello FRBR sono i livelli (cui facevamo già riferimento in cap. 2) che determinano l'esigenza di una descrizione consapevole della nozione di oggetto culturale come entità complessa e stratificata. Sono quelli che ci fanno dire che un'opera ha un contenuto, che determina il livello dell'espressione; che ogni espressione può avere numerose manifestazioni, che corrispondono ai diversi supporti di trasmissione del contenuto; che solo l'item ha lo statuto di unicità.

Ma per ciascuno di questi livelli il tema dei punti di vista eterogeni, della contraddittorietà e dei conflitti di attribuzione, dell'annidamento di ipotesi interpretative e della molteplicità dei contesti rappresentano i problemi da affrontare e che, con il principio del ragionamento per stati evolutivi, possono essere oggetto di modellazione.

In un recente lavoro sul tema della rappresentazione della conoscenza come strumento per dare conto del principio dell'ermeneutica in contesto digitale,[219] abbiamo affrontato il tema della modellazione della conoscenza quando questa sia espressione di punti di vista diversi (conflittuali o semplicemente differenti per il diverso filtro d'osservazione dell'ermeneuta) su uno stesso oggetto d'analisi. Il punto di partenza è stato la definizione del modello interpretativo. Il problema di ricerca è come dare spessore alla descrizione di un artefatto culturale quando questi debba fornire la necessaria espressività a posizioni scientifiche esito di interpretazioni difformi dello stesso fenomeno (o debba documentare il molteplice). E partendo da un caso reale abbiamo cercato di astrarre l'osservazione del dominio per tradurla in un modello concettuale agnostico rispetto agli oggetti osservati.[220]

Il modello dei dati, sul quale torneremo, è quindi stato pensato cercando di qualificare la macro-struttura di osservazione, utilizzando quattro diversi livelli di descrizione (Fig. 8):[221]

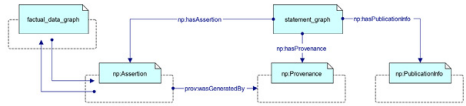

Figura 8 - Il modello dei dati teorizzato nel progetto MIMA (np: è il *namespace* per la *nanopublication*, per cui si veda par. 3.3).
Grafo a cura di Valentina Pasqual. Tratto da http://amsacta.unibo.it/6712/1/AIUCD2021_BOA-versione3A.pdf

- Il livello 0 rappresenta i dati considerati "fattuali" (*factual data*), ossia non questionabili, che vengono considerati come metadati oggettivi;
- Il livello 1, contente le asserzioni (*assertions*), raccoglie i dati considerati potenzialmente contestabili, ovvero le dichiarazioni non universalmente vere o univoche;
- Il livello 2 descrive i dati contestuali (*contexts*) all'asserzione: provenienza, certezza, tipo e criterio dell'interpretazione, fonti, paternità;
- Il livello 3 presenta i dati contestuali relativi alla pubblicazione degli *statements* (*publication*): chi pubblica e quando.

Quelli che qui sono chiamati livelli rispetto all'oggetto d'analisi possono riguardare, in potenza e con valore variabile, ogni entità di Gruppo 1 (in particolare, ovviamente, Expression e Manifestation), consentendo così di stratificare la descrizione dello stesso oggetto in una duplice prospettiva: da un lato i livelli sono le entità di un modello tassonomico (WEMI), dall'altro i livelli sono le diverse sfaccettature attraverso cui osservare ogni singola entità WEMI.

Ecco che da un punto di osservazione di natura bibliografica il modello concettuale di FRBR si allarga a domini non strettamente di natura libraria. E prova il valore di un sistema che ragiona sulla complessità di ogni oggetto culturale come entità stratificata e pluridimensionale.

4.2 Un riuso trasversale: "FRBR all'Opera"

Fino dalle sue origini il modello FRBR ha attirato la comunità, al di là del dominio strettamente bibliografico.[222] Sono varie le

sperimentazioni condotte sul riuso del modello in ambiti diversi da quello delle biblioteche: dall'impiego del modello concettuale per la creazione di ontologie come FABIO[223] o come il già citato modello HiCO, a progetti in cui gli oggetti non sono di natura bibliografica, ma la cui descrizione si può giovare della teorizzazione per livelli ed entità.

Si vuole allora presentare "FRBR all'Opera", per mostrare come il modello sia stato applicato per la descrizione di un contesto non strettamente bibliografico, ovvero la storia del Melodramma Italiano.[224]

Agli inizi degli anni Duemila è stato avviato il progetto RADAMES[225] attraverso il quale si è inteso studiare come organizzare l'informazione all'interno delle collezioni di documenti relativi alle arti performative. L'assunto di partenza è che tale organizzazione non possa prescindere da una rappresentazione delle informazioni documentali, storiche e di contenuto, ovvero dagli strumenti utili per il suo principale utilizzatore: lo storico dello spettacolo. Stante questa necessità, l'adozione del modello FRBR, come linea guida per il disegno del modello descrittivo del sistema informativo sperimentale, è risultata più che naturale. Nel modello descrittivo elaborato dal progetto, a partire dall'impostazione FRBR, sono stati previsti, oltre al catalogo delle diverse tipologie di documenti (testi a stampa e manoscritti, immagini, audio e video ecc.), anche una cronologia degli spettacoli e un repertorio delle opere. Questi ultimi sono pensati per raccogliere l'informazione di natura storica derivata sia dai documenti catalogati che da fonti diverse fruite autonomamente. Da un punto di vista fisico, i due livelli descrivono il piano "immateriale" dei contenuti e degli eventi storici in modo nettamente disgiunto da quello "materiale" dei testimoni del singolo atto performativo. Testimoni che includono non solo i documenti, come i libretti, ma anche tutto il variegato mondo di oggetti che concorrono alla messa in scena dello spettacolo operistico. Per lo storico, in questa dicotomia informativa, resta prevalente la rilevanza dell'informazione consolidata ai primi due livelli, rispetto alla quale, il piano del documento è funzionale. Anche per questo, il modello prevede che cronologia e repertorio possano essere alimentati a partire da fonti esterne al catalogo stesso. In questo modo, viene superata la prospettiva di un modello FRBR applicato all'ambito bibliotecario dove i livelli relativi ai contenuti (opere e relative espressioni)

vengono alimentati a partire dalle edizioni censite nel sottostante catalogo.

L'organizzazione dei contenuti così descritta è pensata in funzione delle necessità di accesso da parte dell'utente finale e non solo al fine di una corretta archiviazione dei materiali. Prendendo come riferimento la comunità di studiosi delle arti performative, il "chi", "cosa" e il "quando" sono le coordinate principali da cui parte la ricerca documentale. La costruzione di una cronologia degli spettacoli, e del corrispondente repertorio, comporta la raccolta delle informazioni relative alla performance e all'opera durante la catalogazione dei documenti. Si tratta quindi di descrivere il documento inteso come testimone di un evento artistico, di individuare le responsabilità degli attori coinvolti nella sua realizzazione, di identificarne i contenuti e di collocarlo nella successione cronologica. Ciascun documento, in funzione della sua natura, sarà così in grado di fornire un variegato insieme di informazioni. Il processo di catalogazione dovrà quindi basarsi necessariamente su una pluralità di fonti per ricostruire l'insieme di elementi informativi utili a descrivere compiutamente l'evento performativo. Certi che, da un punto di vista teorico, nessuna documentazione potrà restituire l'evento nella sua reale complessità, il progetto RADAMES ha potuto verificare come questa raccolta di materiali documentali comporti non solo una diversa modalità di organizzazione dell'informazione, ma anche la stratificazione di diversi livelli descrittivi in fase di catalogazione. Stratificazione che, partendo dal piano materiale del documento, arrivi a descrivere quello dei contenuti passando dalla sua espressione storica: lo spettacolo.

Se i livelli Work, Manifestation e Item sono immediatamente applicabili rispettivamente al repertorio (ad es. l'Othello di Shakespeare), al documento (ad es. il libretto di Otello di Verdi e Boito edito da Ricordi) e all'oggetto materiale (ad es. la copia posseduta dalla Biblioteca della Fondazione Giorgio Cini di Venezia), il livello dell'Expression è rappresentato dalla performance in quanto tale. Nell'applicare il modello al dominio delle arti performative è necessario però tenere in considerazione che l'atto performativo è una espressione effimera dell'opera che non produce una manifestazione stabile della sua esistenza. Lo spettacolo è infatti il risultato dell'interazione dei diversi testi impiegati durante la messa in scena e dell'atto interpretativo che la

realizza. Testo musicale, testo lirico, testo scenico e regia producono, attraverso l'interpretazione vocale e musicale, una espressione dell'opera che resta circoscritta alla durata temporale dello spettacolo. Di conseguenza, i documenti che vengono usati per l'allestimento dello spettacolo d'opera – la partitura, il libretto, i bozzetti scenici, le note di regia ecc. –, devono essere considerati come documenti preparatori. Documenti che, a loro volta, diventeranno manifestazioni di espressioni parziali, o meglio verticali, dell'opera nel suo insieme. A valle della performance, resteranno invece una serie di documenti derivati, come registrazioni audio e video, foto di scena ecc., che sfortunatamente rappresenteranno solo una porzione del contenuto originale dell'espressione prodottasi durante lo spettacolo. Dal punto di vista concettuale, nelle arti performative, l'evento rappresenta l'unica forma in cui le diverse componenti artistiche trovano una loro espressione compiuta. Allo stesso tempo le manifestazioni che ne derivano sono tutte testimoni parziali di un insieme non riproducibile. In sintesi, si può dire che non abbiamo una manifestazione in grado di catturare in toto l'espressione prodotta attraverso la performance. Questo si riflette sul modello descrittivo: il livello dell'Expression potrà raccogliere solo quegli elementi descrittivi della performance che possono essere ricavati dai testimoni parziali della sua esistenza, in primis le informazioni di carattere storico (il "chi", il "quando" e il "dove").

L'articolazione del modello descritta, data l'intrinseca eterogeneità delle entità trattate, comporta una corrispondente articolazione delle relazioni tra le entità. Particolare rilevanza assume il ruolo svolto dagli autori delle diverse componenti che concorrono alla realizzazione della performance. Se si considera il caso specifico del melodramma, a livello di repertorio (Work) abbiamo quantomeno due attribuzioni di responsabilità specifiche: l'autore del testo drammatico e il compositore della partitura musicale. Nel momento in cui si descrive il singolo allestimento (quindi una Expression), ad es. la ripresa dell'*Otello* presso il Teatro alla Scala di Milano nel Carnevale 1889, è opportuno rilevare il ruolo dello scenografo (Giovanni Zuccarelli) e dell'impresario (l'Impresa fratelli Corti & C.) così come i nomi dei cantanti e dell'organico d'orchestra: tutte informazioni riportate nel libretto stampato per l'occasione[226] così come ricavate da fonti secondarie (ad es. la stampa periodica coeva). Più immediata la relazione tra il

libretto (la Manifestation) e l'editore o tra la copia (l'Item) e la biblioteca che lo conserva. D'altro canto, dal punto di vista funzionale, queste specificità di ruolo rendono l'autority file dei nomi una chiave di accesso particolarmente efficace. Ad es., è possibile ricercare tutte le opere edite dall'editore Ricordi oppure conoscere quali ruoli, in opere di Verdi, il tenore Roberto Ramini ha recitato oltre a Cassio e Roderigo nel già menzionato *Otello* (nello specifico, Ismaele nel *Nabucco*). Sarà poi l'utente a decidere se risalire al testo dell'opera per approfondire le caratteristiche del ruolo interpretato o analizzare la cronologia delle interpretazioni per studiare l'evoluzione della carriera dell'interprete. Il modello prevede inoltre la possibilità di indicare rinvii verticali tra i diversi livelli descrittivi e la possibilità di definire legami orizzontali tra entità afferenti al medesimo livello. Ad es., è possibile indicare l'*Othello* di Shakespeare come antecedente bibliografico dell'*Otello* di Arrigo Boito. Dal punto di vista funzionale questo consente all'utente di risalire dai documenti alle opere contenute e di ripercorrerne la tradizione, così come di navigare i contenuti dell'archivio seguendo la coordinata temporale e visualizzare in modo sincrono opere, eventi e documenti.

Il modello descrittivo elaborato dal progetto RADAMES ha trovato una concreta implementazione nell'archivio Corago:[227] il repertorio e archivio di libretti del melodramma italiano dal 1600 al 1900. Il sistema informativo Corago è articolato in tre contenitori: il Repertorio del teatro d'opera, la Cronologia degli spettacoli e la Bibliografia e archivio digitale dei libretti d'opera riprodotti integralmente. Le informazioni storiche e bibliografiche contenute in Corago sono state acquisite dai principali dizionari, dai consueti strumenti bibliografici, dalla consultazione di cataloghi online di biblioteche e di siti Web dedicati al teatro d'opera. La consistenza attuale (al 2021) consta di oltre ventimila opere censite nel repertorio, più di quarantamila eventi nella cronologia e oltre cinquantacinquemila libretti catalogati. Dell'archivio è stata realizzata anche una versione Linked Data ottenuta utilizzando il CIDOC CRM e FRBRoo come ontologie formali di riferimento.[228] L'allestimento della versione LOD ha mostrato l'efficacia delle tecnologie semantiche nel trattare contenuti organizzati secondo un modello concettuale proprietario fortemente orientato ai requisiti di un dominio specialistico come quello del Melodramma. Se dal punto di vista dell'espressività le due ontologie hanno richiesto solo

delle specializzazioni minori per risultare efficaci rispetto al dominio, è dal punto di vista funzionale che l'organizzazione dei dati in formato RDF ha mostrato i suoi limiti. L'estrema analiticità delle due ontologie produce un grafo in cui il singolo concetto viene espresso mediante un numero di triple significativo (si pensi, per esempio, alla data di nascita di un autore). Da un lato l'analiticità consente di circoscrivere la semantica del singolo elemento informativo e quindi di poterlo riapplicare in contesti diversi. Dall'altro però impedisce di fornire una lettura immediata, diciamo sintetica e contestualizzata al dominio, di ciò che il dato descrive. La rappresentazione dell'informazione in questa forma risulta per questo meno efficace dal punto di vista dell'utente finale. Una soluzione è rappresentata dall'adozione di un livello di astrazione attraverso cui applicare le logiche di accesso e rappresentazione del dato specifiche per il singolo dominio.[229] Questo approccio consente di unire l'efficacia funzionale derivante dall'adozione del modello FRBR per descrivere la storia del Melodramma alle potenzialità derivanti dall'accesso al vasto patrimonio informativo rappresentato dall'ecosistema dei Linked Data. In fondo a dire quanto sia importante (ri)pensare alle tecnologie di rappresentazione della conoscenza per fornire un servizio di accesso al sapere per un utente finale.

4.3 Un modello concettuale si trasforma: verso LRMoo

La storia di FRBR è nota: da modello concettuale basato sul paradigma E/R (1998), a ontologia, con il formalismo orientato agli oggetti di FRBRoo (2008), allineato e acquisito nell'altrettanto noto modello CIDOC CRM, fino a RDA (2010), per arrivare a IFLA LRM (2017)[230] e a LRMoo (2020).[231] Attraverso diverse esperienze ed esperimenti.[232]

Come è accaduto nel campo degli archivi, anche in quello delle biblioteche il Semantic Web ha fatto emergere una duplice esigenza: da un lato di unificare modelli relativi ai diversi punti di osservazione di un oggetto – la descrizione dell'oggetto, gli agenti dotati di un qualche ruolo sull'oggetto, gli ulteriori elementi di contesto –, dall'altro il bisogno di riconoscere al ragionamento per triple il ruolo di strumento espressivo di descrizione, al quale

adeguarsi per ripensare al tradizionale sistema di produzione della scheda o dello strumento di corredo. In una prospettiva quindi di integrazione di standard diversi e relativi ad altrettanto diverse chiavi di lettura dello stesso oggetto in un unico modello. Come RiC unifica ISAD, ISAAR, ISDF e ISDIAH, analogamente IFLA LRM vuole integrare FRBR, FRAD e FRSAD. Ecco che quanto abbiamo detto accadere con RiC in campo archivistico, accade con IFLA LRM in campo bibliografico: unificazione degli standard e armonizzazione dei modelli, assieme all'adozione del grafo. Ovvero anche: "vi sono indubbiamente analogie fra i modelli RIC-CM e FRBR-LRM essenzialmente nella struttura adottata per la definizione dei propri concetti, denominata 'entità in relazione'".[233]

Diremo che se la gerarchia costringe ad avvalorare il principio della descrizione separata del soggetto produttore, o il responsabile del contenuto intellettuale, dalla documentazione prodotta o dagli oggetti comunque ascrivibili al *creator*, e anche dalle informazioni necessarie a classificazione e soggettazione, il grafo, pur ribadendo l'autonomia dei tre momenti della descrizione del patrimonio (l'agente, l'oggetto e il contenuto dell'oggetto) riesce a porre ogni entità coinvolta in stretta relazione, avvalorando l'idea di un unico standard che comprenda tutti i possibili contesti di rappresentazione del dato. Saranno le relazioni tipizzate a stabilire il legame fra soggetto e oggetto e ancora fra oggetto e contenuto descrittivo di quell'oggetto.

Ma prima di arrivare a IFLA LRM, una breve riflessione.

FRBRoo, come riconciliazione di un modello di origine bibliografica (FRBR-E/R) con uno di provenienza museale (CIDOC CRM), ha saputo provvedere a un'astrazione capace di raccontare la storia di un complesso digitale eterogeneo per supporto e formato, conciliando le esigenze di descrizione (e concettualizzazione) tipiche dei sistemi bibliografici con quelle in uso in contesti museali. O meglio FRBRoo ha saputo elaborare una proposta che va oltre il dominio strettamente librario. Come si legge nelle specifiche:

> **Libraries and museums** are memory institutions – both strive to preserve cultural heritage objects, and information about such objects, and they often **share the same users**. Besides, the boundary between them is often blurred: libraries hold a number of museum objects and museums hold a number of library objects; the cultural heritage objects preserved in both types of institutions were created in the same cultural context or period, sometimes by the same agents, and they provide evidence of comparable cultural features. It

> seems therefore appropriate to build a **common conceptualisation** of the information gathered by the two types of organisations about cultural heritage.[234]

Quello che ci interessa in questa sede sottolineare è come, con FRBRoo, sia possibile descrivere un oggetto come stratificazione del processo evolutivo che dalla fonte analogica arriva all'oggetto digitale attraverso una certa espressione di una fonte tradotta in potenzialmente molteplici manifestazioni, ma che sia anche possibile stabilire diverse tipologie di interrelazioni, in un dominio non strettamente di descrizione catalografica.

Prendiamo ad es. il recente progetto mythLOD,[235] finalizzato alla trasformazione di un database relazionale – la raccolta Mythologiae, realizzata dal centro di ricerca Framelab dell'Università di Bologna[236] –, in un dataset LOD capace di valorizzare il rapporto fra i metadati descrittivi di un item museale e i frammenti di testi della tradizione (fonti letterarie) capaci di descrivere le componenti iconografiche presenti nell'opera d'arte. In mythLOD la fase di modellazione concettuale ha portato all'impiego del modello dei dati già teorizzato in MIMA (per cui si veda par. 4.1). Come già introdotto, i primi due livelli del modello sono i dati fattuali e le asserzioni, ovvero concetti dipendenti dal dominio di applicazione, mentre il secondo e terzo, ovvero contesti e dati di pubblicazione, sono dichiarazioni trasversalmente applicabili. Per questo motivo, in mythLOD, che documenta un dominio cross-disciplinare, si è deciso di riusare FRBRoo come *backbone ontology* per i primi due livelli applicati a mythLOD (con l'aggiunta di altri elementi da Dublin Core Terms e da Schema data model); FRBRoo ha consentito di gestire la rappresentazione in un sistema che si muove fra contesto museale (gli item interpretati dagli esperti, al centro della collezione) e bibliografico (la relazione individuata dagli esperti tra item e fonti letterarie associate). Nello specifico le opere d'arte sono state modellate come entità appartenenti alla classe efrbroo:F2_Expression e F4_Manifestation_Singleton, mentre le risorse letterarie sono state identificate come entità della classe efrbroo:F1_Work. Inoltre, tra le risorse letterarie presenti nei metadati ("Riscritture Letterarie", "Fonti Classiche", "Fonti Medievali o Moderne", "Riscritture Cinematografiche"), le istanze di fonti classiche sono state oggetto di ulteriore analisi vista la possibilità di formalizzare le citazioni. In particolare, il concetto di efrbroo:F1_Work, è stato scelto per modellare le istanze a una

granularità maggiore tramite il riutilizzo dell'ontologia specializzata HuCit[237] per il concetto di citazione canonica (in particolare hucit:CanonicalCitation) e la possibilità di collegarlo al testo, identificando il frammento specifico attraverso l'uso di URN-CTS.[238]

Tornando a LRM, è indubbio che l'ultima concettualizzazione dell'IFLA introduca una nuova modalità di lettura dell'universo bibliografico,[239] o meglio adegui la rappresentazione delle entità e delle loro relazioni all'approccio ontologico canonico, in una prospettiva di massima astrazione del modello. Ecco che in IFLA LRM, che abbandona la nozione di "record" di FRBR per proporsi come modello per la gestione dei dati riferiti alla "library", compare una superclasse, da cui tutte le entità dipendono, che è la E1 Res (owl:Thing diremo), e viene introdotta la classe E9 Nomen (ripresa da FRAD e FRSAD), per qualificare tanto la RES quanto l'Agent (E6) e ogni entità del dominio, ovvero per fornire un'identità bibliografica attraverso l'associazione fra un'entità e il modo in cui viene designata. Sono ripensate le nozioni di spazio e tempo, ovvero sono introdotte le classi arco temporale e luogo (come entità e non come possibile soggetto di un'entità), mentre scompaiono le classi concetto, oggetto ed evento, dichiarate *deprecated*. La RES è infatti classificata come una qualunque tipologia di entità: un'opera, un manoscritto, una persona dotata di un ruolo, un concetto, un'idea. In continuità con FRBR permangono le entità di Gruppo 1, anche se nuove definizioni le qualificano; il Gruppo 2, gli agenti, è ripensato in struttura gerarchica (una persona singola e un agente collettivo sono le due entità gerarchicamente dipendenti da *agent*); il Gruppo 3, come si diceva, si ridimensiona.

Il Nomen, in particolare, è definito in LRM come una reificazione, rappresenta cioè quella procedura di riduzione a oggetto di un asserto, tale per cui è possibile esprimere proprietà anche relative all'asserto stesso. Così leggiamo nelle specifiche di LRM a proposito dell'entità Nomen:

> Una combinazione arbitraria di segni o simboli non può essere considerata una denominazione o una designazione finché non è associata a qualcosa in qualche contesto. In questo senso, l'entità nomen può essere intesa come la reificazione di una relazione tra un'istanza di res e una stringa. La stringa stessa non costituisce un'istanza dell'entità nomen, ma è modellata come il valore dell'attributo stringa del nomen di un'istanza dell'entità nomen. Due istanze dell'entità nomen possono avere valori perfettamente identici per l'attributo stringa del nomen, tuttavia rimangono distinte finché si riferiscono

> a distinte istanze dell'entità res o hanno valori distinti per uno o più dei loro altri attributi (nonostante si riferiscano alla stessa istanza dell'entità res).[240]

Se ragioniamo in logica RDF, la reificazione consente di usare proprietà che vanno oltre il canonico collegamento soggetto-oggetto, superando la relazione binaria, permettendo quindi di descrivere lo *statement* come una relazione dotata a sua volta di proprietà. È quanto abbiamo già discusso in merito ai grafi nominati, ovvero quella procedura che, permettendo di aggiungere metadati su metadati, consente dichiarazioni finalizzate alla contestualizzazione di affermazioni, come ad es. la menzione di un luogo (ad es. la fonte) da cui una qualche affermazione è stata tratta (per cui si veda par. 3.3).

Se dunque da un lato l'universo bibliografico ha adeguato al modello del Web semantico l'approccio ai principi della descrizione, dall'altro i modelli bibliografici stanno offrendo al Web semantico quello di cui esso ha bisogno per consentire di realizzare appieno gli obiettivi di trasformazione dei dati in oggetti di conoscenza (semantica), tanto computabili dalle macchine quanto capaci di aumentare l'esperienza informativa dell'utente finale.

Con LRMoo, e in generale con CIDOC CRM, ma anche con alcune classi e predicati di RiC-O, si sta lavorando per la creazione dell'ontologia che servirà a rappresentare i dati che confluiranno nella *digital library* del Dipartimento di Filologia dell'Università di Bologna.[241] Questo ambiente raccoglierà i fondi culturali (archivistici e librari) donati alla struttura e conservati presso il Dipartimento. Una prima sperimentazione è ora in fase di test sulla serie dei Quaderni del fondo d'autore Giuseppe Raimondi.[242] Il lavoro sugli archivi di autorità, acquisite le linee guida AIB,[243] ha evidenziato l'importante tema dell'edizione delle carte, riconoscendo che molte delle serie conservate in struttura raccolgono bozze e minute di testi che sono stati pubblicati con varianti in edizioni successive. Questo aspetto in particolare ha portato alla riflessione sulla costruzione di reti di relazioni fra le carte e le opere, fra l'archivio e la biblioteca, ovvero fra risorse che hanno origine dalla stessa idea creativa nativa, che a sua volta ha determinato espressioni diverse (varianti dell'opera), che sono state a loro volta realizzate in differenti manifestazioni.[244] Nella definizione del modello necessario a descrivere la serie dei Quaderni, questo aspetto è stato ritenuto particolarmente importante per la concettualizzazione, premessa necessaria alla

realizzazione dell'ontologia.[245] Ecco che il ragionamento per livelli WEMI ha permesso di dare conto dell'evoluzione del processo creativo dell'intellettuale e di creare quella serie di interconnessioni necessarie a raccontare la storia dell'opera. Un approccio evento centrico, necessario a documentare la storia dei testi nel tempo e attraverso supporti eterogenei.

217 Poi gestite attraverso lo standard *Functional Requirements And Numbering of Authority Records* (FRANAR), https://archive.ifla.org/VII/d4/wg-franar.htm, finalizzato all'estensione del modello di FRBR alle registrazioni d'autorità (punti di accesso alle registrazioni catalografiche). Nel 2001 il gruppo decide di accantonare il progetto di un ISADN (*International Authority Data Number*), e nel 2004 di cambiare la sua denominazione in *Functional Requirements of Authority Records* (FRAR); tra il 2003 e il 2005 pubblica due *drafts*, il secondo dei quali ha per titolo *Functional Requirements for Authority Records*. A conceptual model, che poi diventerà il già citato FRAD (dalla nozione di "record" a quella di "data") nel 2008.

218 Poi gestite con lo standard *Functional Requirements for Subject Authority Records* (FRSAR), che diventerà il già citato FRSAD dal 2010.

219 Marilena Daquino, Valentina Pasqual, Francesca Tomasi, *Knowledge Representation of digital Hermeneutics of archival and literary Sources*, "JLIS.it", 11 (2020), 3, p. 59-76, http://dx.doi.org/10.4403/jlis.it-12642.

220 Il modello originario, realizzato da Valentina Pasqual, si chiama MiMa (Multi-disciplinary Interpretations model on Manuscript Apparatus): "MIMA semantic data model allows to represent the logical and physical structure of an illuminated manuscript, its contents, and scholars' comments on fragments of the manuscript. The project addresses the formal representation of scholars' multidisciplinary analyses performed on the same artefact. A few use cases and examples are taken from a specific manuscript, namely Pellegrino Prisciani's Historiae Ferrarie. Data has been mined from the cycle of lectures called 'Scrivere, rappresentare, conoscere nel rinascimento. Pellegrino Prisciani, un intellettuale eclettico tra la corte e il mondo' and includes statements on philological, art historical, and paleographical aspects related to the illuminated manuscript", https://mima-data-model.github.io/mima-documentation (accesso Web con interfaccia utente). Il data model e tutta la documentazione sono su Github, https://github.com/mima-data-model/mima-documentation. Si veda anche Valentina Pasqual, Marilena Daquino, Francesca Tomasi, *MIMA: A Data Model to Represent Multi-Disciplinary Analysis on Manuscripts. Use Case on*

Pellegrino Prisciani's Historiae Ferrariae, in *AIUCD 2021, Book of extended abstracts*, Quaderni di Umanistica Digitale, p. 560-566, https://doi.org/10.6092/unibo/amsacta/6712.

221 Livelli rappresentati formalmente con grafi nominati seguendo il modello nanopublication (par. 3.3).

222 Si veda Karen Coyle, *FRBR Before and After*, Chicago, ALA, 2016; trad. it. *FRBR prima e dopo*, a cura di Lucia Sardo, AIB, 2017.

223 FRBR-aligned Bibliographic Ontology (FaBiO), http://www.sparontologies.net/ontologies/fabio.

224 Si ringrazia Paolo Bonora per le pagine che seguono (par. 4.2).

225 Lorenzo Bianconi, Angelo Pompilio, Gennaro Pagannone, *RADAMES: prototipo d'un repertorio e archivio digitale per il melodramma*, "Il Saggiatore Musicale", 11 (2004), 2, p. 345-394.

226 Si veda la scheda dedicata all'*Otello* all'URI http://corago.unibo.it/resource/DOCUMENTI/0001081388. Copia digitale del libretto, https://archive.org/details/otellodrammaliri00verd_1/mode/2up.

227 Corago, sito ufficiale, http://corago.unibo.it. Il progetto è curato da Angelo Pompilio presso l'Università di Bologna.

228 Paolo Bonora, Angelo Pompilio, *Corago in LOD. The debut of an Opera repository into the Linked Data arena*, "JLIS.It", 12 (2021), 2, p. 54–72, https://doi.org/10.4403/jlis.it-12699.

229 Paolo Bonora, Angelo Pompilio, *"Osservate, leggete con me". Risorse LOD per la storia del melodramma: una prospettiva funzionale di rappresentazione*, in *AIUCD 2019*, cit, p. 202-207.

230 La versione attuale, 2017, è la già citata di Pat Riva, Patrick Le Bœuf, Maja Žumer, *IFLA Library Reference Model*, cit. Si veda poi in particolare Pat Riva, *Il nuovo modello concettuale dell'Universo bibliografico: FRBR Library Reference Model*, "AIB studi", 56 (2016), 2, p. 265-275. In generale si possono consultare le pagine dell'IFLA dedicate al modello, https://www.ifla.org/publications/node/11412.

231 Library Reference Model - Object-oriented (LRMoo). Ancora in stato di bozza: http://www.cidoc-crm.org/ModelVersion/lrmoo-f.k.a.-frbroo-v.0.6. Tutte le versioni in https://www.cidoc-crm.org/frbroo/fm_releases.

232 Si veda, ad es., Paolo Ciccarese, Silvio Peroni, *The Essential FRBR in OWL2 DL Ontology (FRBR)*, "an expression in OWL 2 DL of the basic concepts and relations described in the IFLA report on the Functional Requirements for Bibliographic Records (FRBR), also described in Ian Davis's RDF vocabulary", 2018, http://purl.org/spar/frbr.

233 Susanna Peruginelli, et al., *RDA e archivi: ricerca di un raccordo tra mondi diversi*, "JLIS.it", 9 (2018), 1, p. 137-147, https://dx.doi.org/10.4403/jlis.it-12403.

234 Chryssoula Bekiari, Martin Doerr, Patrick Le Bœuf, Pat Riva, *Definition of FRBROO. A Conceptual Model for Bibliographic Information in Object-Oriented Formalism*, cit. Grassetto nostro.

235 Il progetto, a cura di Valentina Pasqual, è online all'indirizzo https://dharc-org.github.io/mythlod. Ma si veda anche Valentina Pasqual, Francesca Tomasi, mythlod/1.0, version 1.0, 2021, http://doi.org/10.5281/zenodo.4639821.
236 Il progetto Mythologiae è consultabile all'indirizzo https://mythologiae.unibo.it.
237 Humanities Citation (HuCit), http://purl.org/net/hucit.
238 URN è il sistema di identificazione univoca dei nomi (Uniform Resource Names) per Canonical Text Services (CTS), http://cite-architecture.github.io/ctsurn_spec.
239 Si può vedere il grafo, che mostra una *overview* di entità e relazioni di LRM, nelle specifiche di progetto, https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf, p. 86.
240 Pat Riva, Patrick Le Bœuf, Maja Žumer, *IFLA Library Reference Model*, trad. it., cit. , p. 31.
241 La *digital library* del Ficlit è consultabile all'indirizzo https://dl.ficlit.unibo.it. Si tratta di un ambiente basato sul *repository* semantico Omeka S, con esposizione delle immagini in IIIF, e visualizzazione delle immagini stesse attraverso il plug-in Mirador. I dati che confluiranno nella *digital library* provengono da tipologie documentarie analogiche eterogenee (principalmente fondi archivistici e bibliografici, a stampa e manoscritti). La *digital library* sarà quindi territorio utile per ragionare sulle prospettive di dialogo di modelli diversi nel dominio MAB/(GLAM). In particolare si sta lavorando sulla documentazione archivistica per elaborare un modello, riusando anche standard non prettamente archivistici (CIDOC CRM in particolare e quindi FRBRoo/LRMoo). La comunità archivistica d'altronde è predisposta al riuso di modelli di derivazione bibliografica. Si veda, ad es., Federico Valacchi, *Things in the World. The integration process of archival descriptions in intercultural systems*, "JLIS.it", 7 (2016), 2, p. 331-367, https://dx.doi.org/10.4403/jlis.it-11529.
242 Giuseppe Raimondi (Bologna 18 luglio 1898 - 3 agosto 1985). Il fondo consta di: 460 Quaderni di appunti, racconti e altri manoscritti letterari (organizzati in 3 sottoserie); 479 articoli redatti dall'autore per la pubblicazione su quotidiani (di ogni articolo si conserva generalmente la copia manoscritta, quella dattiloscritta e quella a stampa); 828 cartelline, ordinate alfabeticamente in base al mittente, contenenti corrispondenza, ordinate alfabeticamente per mittente; 13 Album che raccolgono tutto il materiale riguardante i più importanti corrispondenti di Raimondi (Apollinaire, Ungaretti, De Pisis, ecc.). Ogni album contiene tipologie documentarie diverse, come cartoline, lettere, fotografie, opuscoli, immagini, ritagli di giornali, disegni originali, fotocopie, ecc. L'elenco delle serie Corrispondenza e Album è pubblicato in *Giuseppe Raimondi: carte, libri, dialoghi intellettuali*, Bologna, Patron, 1998, p. 179-311. Tutto il materiale è stato descritto attraverso ISBD ed è confluito in OPAC SBN. Si avverte

dunque l'esigenza di una rilettura di tali descrizioni alla luce della dimensione archivistica e l'ontologia servirà esattamente a ricreare i legami fra i livelli della descrizione.

243 *Linee guida sul trattamento dei fondi personali*, a cura della Commissione nazionale biblioteche speciali, archivi e biblioteche d'autore, versione 15.1, 2019, https://www.aib.it/wp-content/uploads/2020/01/15.1_Linee-Guida-fondi-personali-def-1-ULTIMA-VERSIONE.pdf.

244 Su questo tema si veda Francesca Tomasi, *L'edizione delle carte degli scrittori: l'archivio letterario incontra il (nativo) digitale*, "Biblioteche oggi Trends", 7 (2021), 1, p. 81-91, http://dx.doi.org/10.3302/2421-3810-202101-081-1.

245 Una prima proposta si legge in Francesca Giovannetti, Francesca Tomasi, *The Linked Finding Aid as a Platform for Textual Research: The Case Study of the Giuseppe Raimondi Archive*, in *Linked Archives 2021*, in *Proceedings of Linked Archives*, International Workshop 2021, co-located with 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021), "CEUR Workshop Proceedings", 3019 (2021), p. 104-113, http://ceur-ws.org/Vol-3019/LinkedArchives_2021_paper_3.pdf.

5. ORGANIZZARE LA CONOSCENZA CON I LOD

Exposing library and archival content as linked data represents a different working style, but the end result is the same – making the content of our collections available for use and understanding.[246]

5.1 Cosa significa organizzazione della conoscenza?

Per dirla con le parole di Claudio Gnoli:

> L'organizzazione della conoscenza è la disciplina dedicata all'insieme delle tecniche con cui la conoscenza può essere ordinata in modi utili per la sua consultazione e il suo utilizzo. Essa comprende tutti gli approcci semantici all'informazione (schemi di classificazione, tesauri, intestazioni di soggetto, terminologia, tassonomie, ontologie e linguaggi per la rappresentazione della conoscenza) e la loro applicazione nella documentazione, nella ricerca, nell'organizzazione del lavoro e nella società in genere.[247]

Parliamo dunque di tecniche di gestione del sapere, applicabili attraverso tipi diversi di strumenti semantici. Ormai sappiamo che l'informazione, quando adeguatamente strutturata attraverso modelli di rappresentazione, diventa conoscenza; gli strumenti semantici sono quelli che garantiscono alla conoscenza di essere organizzata, in modi e forme diversi, facilitando quindi l'accesso al sapere da parte di un'utenza profilata, ovvero caratterizzata da esigenze diverse di fruizione della conoscenza, e quindi che necessita di servizi differenziati – lo studioso esperto di dominio, il catalogatore che deve descrivere i dati, lo studente che se ne deve servire a scopi di apprendimento, il pubblico che vuole conoscere qualcosa di nuovo attraverso l'esplorazione dei dati.

All'avvio del corso di Knowledge Organization and Cultural Heritage[248] attivo presso la laurea magistrale internazionale di Digital Humanities and Digital Knowledge (DHDK),[249] in genere viene introdotta la definizione di KO con questa nota citazione:

> Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, databases, archives etc. These activities are done by librarians, archivists, subject specialists as well as by computer algorithms. KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations and concepts. There exist different historical and theoretical approaches to and theories about KO, which are related to different views of knowledge, cognition, language, and social organization. Each of these approaches tends to answer the question: "What is knowledge organization?" differently. LIS professionals have often concentrated on applying new technology and standards, and may not have seen their work as involving interpretation and analysis of meaning.[250]

Ecco che la conoscenza, ovvero anche quanto emerge dai documenti intesi come oggetti culturali in senso ampio, deve fare i conti con due concetti che, nell'approccio tradizionalmente monodimensionale dell'atto di descrizione degli artefatti, mancano: l'interpretazione e l'analisi del significato. Perché non c'è conoscenza senza analisi e interpretazione. Quello che nel contesto della LIS è percepito come estraneo (anche se estraneo non è), diventa con le DH, attraverso il Web semantico, il cuore dell'approccio critico alle fonti storiche o anche ai dati del patrimonio culturale che si può tradurre in un asserto: la conoscenza come esito dell'ermeneutica. L'organizzazione della conoscenza nelle DH trova il suo primato esattamente nella modellazione dei dati, in prima battuta a livello concettuale, l'attività che sta a fondamento di ogni processo finalizzato all'analisi di un dominio in vista della sua manipolazione computazionale.[251]

Ma prima di poterla organizzare, dobbiamo darle un significato e quindi chiederci: cosa significa conoscenza? Se volessimo recuperare la già citata piramide DIKW,[252] diremo che la conoscenza è un'informazione dotata di contesto. E il contesto, come abbiamo detto, è il corredo meta e para testuale che attribuisce significato all'informazione. Abbiamo già avuto modo di parlarne nei capitoli precedenti. Qui ci interessa capire perché si parla di conoscenza e perché va organizzata. Nel contesto del Web

semantico, il concetto di conoscenza è legato, evidentemente, a quello di semantica. Ovvero la semantica è la componente necessaria a rendere la macchina capace di computare il sapere. Potremmo dire che l'ontologia è lo strumento che trasforma un dato, che è diventato informazione perché arricchito di struttura, in conoscenza. Se la struttura, possiamo dire, stabilisce la sintassi, la semantica determina il significato. Come abbiamo già detto, un dato marcato con un linguaggio di markup è un'informazione, perché l'annotazione dichiarativa si posiziona ancora a livello di sintassi. Quando quello stesso dato sia associato a un'ontologia, che lo descrive come una classe o una proprietà, o anche una relazione o un'istanza, allora quel dato diventa conoscenza. Questo è il motivo per cui si parla di grafo della conoscenza, che abbiamo già altrove chiamato *knowledge graph*, capace di determinare la base di conoscenza (*knowledge base*), ovvero i dati, assieme ai modelli concettuali e agli strumenti per potervi accedere (le query). Questo è anche il motivo per cui abbiamo chiamato l'ambiente Web di fruizione di un'edizione digitale semantica un *knowledge site*.[253] Perché dietro alla rappresentazione del testo e la sua ridefinizione in dataset sta sempre un modello concettuale, espresso attraverso un'ontologia, che rende il grafo computabile: un'informazione dotata di contesto.

Organizzare la conoscenza così intesa, ovvero come informazione dotata di contesto che stabilisce la semantica, significa individuare strumenti che valorizzino la capacità di trasmissione del sapere.

Se le ontologie, evoluzione diremo di tassonomie e thesauri, sono tipicamente pensate come strumenti di organizzazione della conoscenza, perché strutturano un dominio per classi, sottoclassi, proprietà e istanze, il nuovo paradigma dei LOD può essere pensato come modalità di utilizzo delle ontologie per la rappresentazione della conoscenza. Le DH contribuiscono fornendo nuove metodologie per la costruzione di *Knowledge Organization Systems* (KOS) attraverso i LOD.[254]

Riprendendo le parole di Cristina Pattuelli:

> In a recent interview, Bernhard Haslhofer (Blumauer 2010) suggested that linked data technology represents a natural extension of the library practice of building knowledge organization tools, including metadata, controlled vocabularies, and identifiers. According to Haslhofer, linked data can be seen as "a natural technical evolution step in information organization" (para 4). For their part, libraries and other cultural heritage institutions have the potential to make a significant contribution to the linked data context by

> sharing their extensive collections of high quality metadata and authority data, providing a "backbone of trust" (Hannemann & Kett, 2010, 2). Despite its promise, much of the work is still in its early stages and a number of **challenges** need to be addressed before libraries, museums, and archives are able to take full advantage of **linked data consumption** (Coyle, 2010; Byrne & Goddard, 2010).[255]

Se quindi tanto è stato fatto dagli istituti culturali, tanto ancora rimane da fare.[256] E se nel 2010 ci si lamentava che ancora i LOD dovessero affrontare un certo numero di "challenges" prima che il GLAM potesse trarre vantaggio da questo nuovo sistema di organizzazione delle informazioni, ancora nel 2020 la strada è appena avviata. In questa prospettiva ecco che il ruolo delle DH è proprio quello di coniugare sapientemente KOS e LOD per valorizzare un autentico accesso alla conoscenza. Ovvero usare i LOD come strumento necessario a creare ambienti a base semantica, e usare quei dati come fonte di arricchimento per l'utente finale. Se tanto infatti si è fatto per la creazione di datasets molto ancora rimane da fare sul fronte del vero impiego dei LOD come ausilio alla conoscenza. E questo significa tanto mettere in campo un approccio analitico ai dati, quanto far valere il principio dell'interpretazione di quei dati, come espressione soggettiva di chi quei dati ha descritto, per fare in modo che esprimano un sapere potenzialmente latente, veicolato dai contesti. Ma senza dimenticare che questi dati devono essere fruiti, ovvero devono essere garantiti servizi per l'accesso ai dati e quindi per la loro valorizzazione, anche a fronte di esigenze diverse di fruizione del sapere ("linked data consumption" così come qualificato nella citazione di poco sopra).

Pensiamo a un caso esemplare, come può essere l'archivio culturale, che potremmo tradurre nel concetto di fondo d'autore (o d'autorità) e che abbiamo già introdotto (si veda par. 4.3). Archivi di letterati, intellettuali, artisti che hanno lasciato alle istituzioni deputate alla conservazione del patrimonio tanto la documentazione che si è sedimentata come esito della loro attività (dalle lettere agli estratti, dai quaderni di appunti alle foto, da ritagli di giornale ai fogli sparsi), quanto il fondo librario, o magari collezioni di artefatti diversi (quadri, sculture e oggetti culturali) reperibili in altrettanto diversi istituti di conservazione e quindi consultabili attraverso differenti strumenti di corredo. Si tratta di uno degli esempi più trasversali nel dominio del MAB. E questa

trasversalità va analizzata in un approccio che traduca l'interpretazione dello studioso alla luce del contesto in cui i dati sono inseriti, ma anche riconoscendo i diversi livelli della rappresentazione. Organizzare una conoscenza che per sua natura nasce in contesti culturali diversi significa fare uno sforzo di allineamento di modelli tradizionalmente differenti sul piano degli standard di riferimento, ma dove i dati si riferiscono a concetti tipicamente in mutua relazione perché interconnessi da un qualche punto di vista.

Se organizzare la conoscenza significa allora creare nuovi vocabolari capaci di descrive lo spessore semantico dei dati, il tema dell'organizzazione si riflette sulla capacità del fruitore di poter accedere al sapere, a prescindere da dove le risorse, portatrici di quel sapere, sono conservate e dallo standard che è stato utilizzato per descriverle.

Si vuole ragionare sull'importanza di poter disporre di *digital libraries* di materiali eterogenei disseminati, capaci di raccogliere il catalogo dei materiali archivistici, bibliografici e museali e di erogare nuovi servizi di accesso e fruizione attraverso cui aumentare la capacità dell'utente finale di ampliare il proprio orizzonte conoscitivo.[257] Se il modello dell'Open Archival Information System (OAIS) determina le coordinate del modello teorico (dall'architettura generale alla suddivisione degli *information packages* sulla base del loro ruolo, ovvero *submission* SIP, *archiving* AIP e *dissemination* DIP),[258] lo sforzo di organizzare oggetti eterogeni, descritti attraverso standard differenti, ma capaci di esprimere a livello concettuale le interconnessioni fra dati, spinge a ragionare su come valorizzare un sistema di relazioni trasversali fra entità. E corpora eterogenei di dati culturali danno spazio a una molteplicità di narrazioni, determinate dai diversi percorsi attraverso cui esplorare il patrimonio culturale. Ecco che il tema dell'organizzazione della conoscenza va oltre i suoi tradizionali confini per diventare strumento utile a comprendere i dati attraverso la narrazione che di quei dati è stata fatta dall'interprete, che ha individuato i contesti necessari a garantirne l'espressività, ma ha anche scelto le relazioni attraverso cui raccontare i legami fra gli oggetti del patrimonio stesso. Ed è il fruitore che potrà tanto accedere ai dati quanto alla narrazione creata da chi quei dati ha studiato. Senza ovviamente escludere un potenziale contributo attivo di un lettore sapiente. E auspicando la nascita di narrazioni

non preordinate e quindi inaspettate.

5.2 Dati aperti e interconnessi

Abbiamo ormai chiarito che ci troviamo nell'era del cosiddetto Web of data. Il dato è l'elemento atomico, il primo strumento necessario alla costruzione della conoscenza. Big data, open data e linked data sono le parole chiave di un'epoca che individua nel frammento la base di partenza per la costruzione di entità. Entità necessariamente identificate in modo univoco e poste in relazione attraverso quei predicati che stabiliscono la semantica della rete di conoscenza.

Il dato è canonicamente computabile sulla base della struttura che gli assegniamo. Abbiamo già parlato di dati non strutturati, come è il testo pieno di un documento, e di dati strutturati, come i dati che individuano nel database il modello di riferimento. E poi ci sono i dati semi-strutturati, tipicamente quelli che sono esito dell'impiego di linguaggi di markup (ad es. XML), quei documenti, diremo, in cui la struttura parziale è determinata dall'impiego di marcatori come strumenti sintattici per l'annotazione.

Quando pensiamo ai LOD da un lato ragioniamo su dati strutturati, che tipicamente possono essere rappresentati in un database a grafo che fornisce il livello astratto di concettualizzazione; dall'altro, potendo esportare i nostri LOD in uno dei diversi formati di serializzazione esistenti, pensiamo al dataset come una collezione semi-strutturata (RDF), in cui però la semantica non è stabilita dalle etichette per l'annotazione come in un comune file XML, ma dall'ontologia, che rappresenta il modello concettuale al pari di quanto avviene in un database nella prima fase di progettazione a livello astratto (tipicamente modello UML o E/R).

Ecco che i LOD si esprimono in una fondamentale costante interazione fra il dato e il metadato, fra l'unità atomica e il suo ruolo ontologico (appartenenza a una classe o dichiarazione di relazione con un oggetto).

Quanto ci preme qui enfatizzare è che la modellazione dei dati, che qualificano un dominio, è un'azione che riflette tanto il punto di vista dell'osservatore quanto gli obiettivi della computazione. Ecco allora che sarà necessario innanzitutto capire cosa significhi

produrre dati interconnessi, e quindi quali siano le problematiche di questo processo di costruzione di reti di conoscenza.

Proviamo dunque a riflettere sul *workflow* del processo di creazione di LOD. Tipicamente quando si voglia pubblicare un dataset in LOD ci si potrà confrontare con almeno due possibili situazioni: una conversione di dati che partano da un qualche modello di rappresentazione (database relazionale o file XML descritti attraverso una qualche grammatica), o la creazione ex novo di un dataset. Le linee guida o buone pratiche per la produzione di LOD (o per la conversione di esistenti datasets) ormai sono numerose.[259] E certamente il *workflow* di questo percorso vede sempre come punto di partenza l'analisi dei dati di input, perché è l'osservazione (critica) dei fenomeni (e delle entità) potenzialmente veicolati dalla collezione di dati che permette di modellizzare in vista della rappresentazione in RDF. Ed è il modello che consente di stabilire il set delle relazioni necessarie a documentare il dato e a esprimere queste relazioni sulla base delle ontologie, o più genericamente dei vocabolari, più appropriate tanto a descrivere il dominio quanto a rispondere ai bisogni della ricerca. Bisogni, che come emergono dalle cosiddette *competency questions*, necessarie in fase di analisi dei dati per avviare il processo di modellazione.

Ora, il caso più comune di produzione di LOD è quello della conversione. Dati già rappresentati attraverso un qualche sistema formale che devono essere ripensati in prospettiva semantica. E in questo caso il processo non sarà comunque né lineare e nemmeno univoco, ma certamente richiederà un percorso iterativo e, di nuovo, l'assunzione di un punto di vista.

Prendiamo quattro esempi di *workflow* relativi a quattro casi diversi, che facciano emergere le criticità del processo. L'esperienza condotta nel già citato progetto Zeri&LODe (per cui si veda par. 2.3) documenta il caso di un dataset di partenza (archivio di fotografie metadatate e correlate alle opere d'arte, soggetto delle fotografie) altamente strutturato nella sua forma nativa (database relazionale), con un controllo sul valore dei descrittori (attraverso l'impiego di vocabolari controllati). Il cuore del processo è qui stato il riuso di ontologie esistenti,[260] ma anche la trasformazione degli Schemi di metadazione dell'ICCD (scheda F[261] per le fotografie e scheda OA[262] per le opere d'arte) in ontologie (F entry[263] e OA entry[264]) e il *mapping* con CIDOC CRM.

In un altro recente progetto, mythLOD (per cui si veda par. 4.3),

la criticità principale è stata invece rappresentata dall'assenza di controllo sui valori delle etichette descrittive. Questo ha significato uno sforzo importante nella fase di cosiddetto *data cleaning* (anche chiamato altrove "bonifica"[265]), ovvero di normalizzazione dei valori espressi in linguaggio naturale (nomi di persona, luoghi, date, citazioni, parole chiave) che naturalmente soffrono di ambiguità, indeterminatezza, ma anche di errori o di mancata standardizzazione, affinché possano essere rappresentati in un sistema formale, identificati univocamente attraverso adeguati URI e "riconciliati" ai sistemi di controllo d'autorità (VIAF e Wikidata in particolare).

Nel già descritto caso del quaderno di Paolo Bufalini (per cui si veda par. 3.3) il *workflow* ha la particolarità di documentare un'attività condotta sul full-text di un'edizione, che ha richiesto in prima battuta un arricchimento delle entità, rappresentate da marcatori XML/TEI, per renderle convertibili in un dataset a base LOD. In questo caso il grafo della conoscenza è collegato al testo dell'edizione attraverso puntatori, che sono espressi sotto forma di frammenti di URI nel file XML/TEI, necessari a costruire le relazioni nella *knowledge base*.[266] Lavorare sul full-text mantenendo un approccio data-centrico nella rappresentazione della conoscenza è un'altra esigenza sentita nel contesto delle edizioni digitali, soprattutto quando ci si voglia muovere in un contesto semantico.

E veniamo allora all'ultimo caso, quello rappresentato dal recente lavoro condotto per l'edizione digitale delle opere di Aldo Moro.[267] In questo progetto l'idea è stata di utilizzare RDF nella sua serializzazione RDFa per annotare le triple direttamente nel testo (*embedding* della semantica nell'HTML) e valorizzare le entità ritenute interessanti: persone, luoghi, date, citazioni dirette, a cui si aggiungono ruoli (per le persone), temi (per il contenuto) e tipologie (per i documenti).[268] L'output dei file è in vari formati, fra cui XML/TEI per garantire la massima interoperabilità con altre edizioni esistenti.[269]

Ma affinché i LOD siano uno strumento utile anche per la ricerca scientifica, ovvero che rispondano a reali bisogni nelle humanities, vale a dire che i dati siano in grado di rispondere a veri quesiti di ricerca, qualche passo in più andrà fatto. Abbiamo già detto del tema dell'atto interpretativo, del ragionamento per livelli e dell'aggiunta dei necessari contesti ai dati. Diciamo che i LOD rappresentano quell'anello di congiunzione fra archivi e biblioteche

in un sistema che dovrebbe facilitare il dialogo fra istituzioni culturali che adottano standard di descrizione differenti, ma che possono dialogare in un nuovo sistema interconnesso. Questo significa, come abbiamo già avuto modo di dire, che l'arricchimento non è il solo controllo di autorità ma la creazione di collegamenti fra fonti eterogenee, concettualmente correlate. Ecco che lo sforzo che le istituzioni culturali devono fare è quello di costruire ponti fra le risorse mettendo a dialogo il patrimonio culturale in un circuito virtuoso in cui sia possibile spostarsi fra dati archivistici, bibliografici e museali, integrando servizi capaci di consentire al fruitore di costruire percorsi di lettura.

Prendiamo un caso particolare, come quello delle edizioni, dove sicuramente la semantica del grafo derivante varierà a seconda di una serie di parametri (per cui si veda anche par. 2.4): il livello di osservazione dell'oggetto che si sta descrivendo (WEMI), il contesto in cui un'entità occorre (una persona dotata, ad es., di un ruolo che varia al variare del luogo della sua menzione) e soprattutto il punto di vista di chi modella il testo, che porta nel modello saperi e conoscenze derivanti dal suo background culturale e sociale; saperi e conoscenze che sono capaci di stabilire una certa rappresentazione del dominio osservato proprio perché espressione di un punto di vista: l'osservazione del filologo, rispetto a quella dello storico o ancora del teorico della letteratura o del linguista, tanto per prendere alcuni esempi, determina sicuramente una diversa modellazione dei dati.

Date queste riflessioni è evidente allora che produrre dati in LOD non è solo una modalità alternativa di pubblicazione, ma è l'esigenza di valorizzare le entità espresse da cataloghi, inventari e repertori cercando di andare oltre la pura classificazione delle entità canoniche, quelle che rispondono alle domande "chi", "cosa", "quando" e "dove", ma che usino le nozioni di persona, soggetto, data e luogo come strumento per costruire narrazioni. Diremo che i LOD sono a tutti gli effetti un nuovo strumento di corredo, ovvero un nuovo modo per esplorare la capacità di raccontarsi da parte degli oggetti culturali, che sono tanto le risorse conservate presso le nostre istituzioni quanto il prodotto della ricerca di quegli studiosi che, partendo da quelle stesse fonti digitali, ne hanno voluto raccontare la storia attraverso progetti capaci di rispondere a bisogni conoscitivi nuovi.

L'obiettivo verso il quale è necessario dunque investire è di

rendere i LOD uno strumento, e questo significa offrire servizi.

5.3 Visualizzare i dati. Le nuove frontiere orientate all'utente

Se tanta letteratura è stata prodotta sulla realizzazione di LOD, ancora tanta strada c'è da fare nella realizzazione di applicazioni che siano in grado di utilizzare in modo sapiente quei dati per restituirli all'utente sotto forma di nuova conoscenza.

Partiamo dal presupposto che i LOD non sono pensati originariamente per l'utente, ma per l'elaborazione da parte della macchina. È però evidente che per poter usare questi dati è necessario che essi vengano "esposti" in un modo che vada oltre l'asserto sotto forma di tripla RDF. Il linguaggio di interrogazione per RDF, ovvero SPARQL, non è un linguaggio per tutti, in quanto richiede una certa competenza per poter formulare un'interrogazione; SPARQL non può dunque essere pensato come un sistema di interrogazione per tutte le tipologie di utenti. A questo si aggiunge la complessità di effettuare interrogazioni complesse direttamente su un *endpoint* SPARQL (problema tecnico), che si somma all'esigenza di fornire interfacce di accesso amichevoli per l'utente finale (problema teorico).

Le prime realizzazioni di visualizzazioni per LOD, che iniziano a circolare attorno al 2010, richiamano il canonico approccio della scheda catalografica, quello stesso approccio, ad es., che usa DBpedia. Soggetto come intestazione, lista dei predicati a sinistra e oggetto a destra (Fig. 9). Ecco che al listato di triple di una query SPARQL si affianca non solo la visualizzazione tabellare (ad es. LodView, Fig. 10), ma anche la manipolazione interattiva di un grafo con i nodi interconnessi e le relazioni semantiche esplicite (ad es. LodLive, Fig. 11).

È evidente che affinché la visualizzazione delle triple serva davvero all'utente come mezzo per acquisire quella conoscenza che i dati veicolano, altri strumenti saranno necessari. Perché è vero che visualizzare è un servizio per l'utente, che può utilizzare un sistema di accesso amichevole – ovvero quello a cui ci hanno abituato gli strumenti informatici, dai sistemi operativi ai programmi di videoscrittura, fino ai browser per il Web –, ma visualizzare è anche individuare strategie che valorizzino la cosiddetta *user experience*,

ovvero è sposare quell'approccio olistico a dati, servizi e strumenti, con lo scopo di agevolare l'utente nella scoperta di qualcosa di nuovo. Senza dimenticare che la visualizzazione grafica è il primo sistema di controllo della bontà del codice e quindi di verifica della presenza di potenziali errori.

Facciamo un passo indietro.

Il tema di come organizzare la conoscenza ha subito una rapida impennata con l'avvento del Web. Ciò che Ranganathan aveva ipotizzato con la sua CC diventa con il Web uno strumento per descrivere risorse attraverso le faccette come modello di organizzazione della conoscenza;[270] ma soprattutto le faccette sono pensate per riconfigurare i metadati del Web, come ad es. avviene con il Dublin Core e i suoi descrittori, che si possono definire un veicolo di categorizzazione delle risorse, capace di stabilire e veicolare i diversi punti di osservazione di un oggetto digitale.

Ecco che in termini di architettura dell'informazione, descrivere gli item di una collezione di oggetti culturali attraverso un set di descrittori stabilisce i criteri di selezione dei dati, azione necessaria a filtrare i risultati.

Sono ormai numerosissime le collezioni di oggetti culturali che utilizzano categorie concettuali come strumento di esplorazione di un corpus: persone, luoghi, date, tipologie documentare, ma anche licenze o diritti, istituti di conservazione, giusto per stabilire qualche tipologia, diventano gli strumenti che, se da un lato servono a organizzare l'informazione dall'altro sono mezzi di esplorazione della conoscenza. E questo accade soprattutto quando le faccette siano il risultato dell'uso di classi e predicati di un'ontologia come canali di navigazione della raccolta.[271] Ecco che, di nuovo, le domande "chi", "cosa", "quando" e "dove" diventano le chiavi primarie attraverso cui esplorare un dataset. A cui dovremmo aggiungere il "perché", ovvero la dimensione dell'interpretazione, che significa, come abbiamo già avuto modo di dire, documentare i contesti necessari a valorizzare la dimensione scientifica (*scholarly*) dei dati, a garanzia di qualità del contenuto e quindi di affidabilità del messaggio (come l'attribuzione di paternità della descrizione, la fonte di riferimento per l'asserzione e la certezza della dichiarazione).

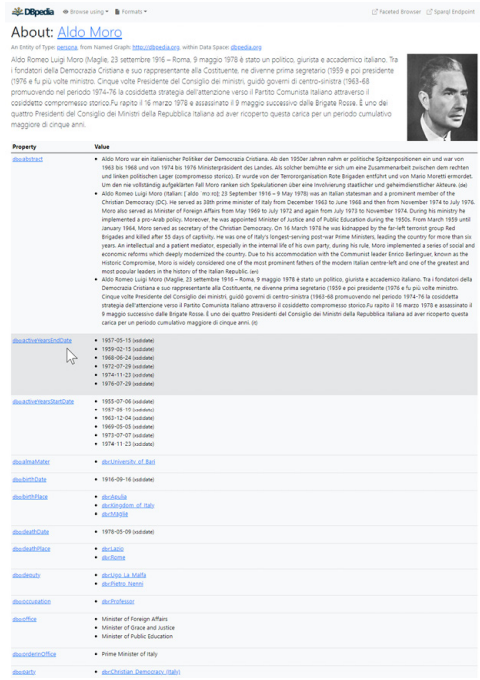
DBpedia
Browse using
Formats
Faceted Browser
Sparql Endpoint
About: Aldo Moro
An Entity of Type: persona, from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org
Aldo Romeo Luigi Moro (Maglie, 23 settembre 1916 – Roma, 9 maggio 1978 è stato un politico, giurista e accademico italiano. Tra i fondatori della Democrazia Cristiana e suo rappresentante alla Costituente, ne divenne prima segretario (1959 e poi presidente (1976 e fu più volte ministro. Cinque volte Presidente del Consiglio dei ministri, guidò governi di centro-sinistra (1963-68 promuovendo nel periodo 1974-76 la cosiddetta strategia dell'attenzione verso il Partito Comunista italiano attraverso il cosiddetto compromesso storico.Fu rapito il 16 marzo 1978 e assassinato il 9 maggio successivo dalle Brigate Rosse. È uno dei quattro Presidenti del Consiglio dei Ministri della Repubblica italiana ad aver ricoperto questa carica per un periodo cumulativo maggiore di cinque anni.
Property
Value
dbo:abstract
Aldo Moro war ein italienischer Politiker der Democrazia Cristiana. Ab den 1950er Jahren nahm er politische Spitzenpositionen ein und war von 1963 bis 1968 und von 1974 bis 1976 Ministerpräsident des Landes. Als solcher bemühte er sich um eine Zusammenarbeit zwischen dem rechten und linken politischen Lager (compromesso storico). Er wurde von der Terrororganisation Rote Brigaden entführt und von Mario Moretti ermordet. Um den nie vollständig aufgeklärten Fall Moro ranken sich Spekulationen über eine Involvierung staatlicher und geheimdienstlicher Akteure. (de)
Aldo Romeo Luigi Moro (Italian: [aldo mo ro]; 23 September 1916 – 9 May 1978) was an Italian statesman and a prominent member of the Christian Democracy (DC). He served as 38th prime minister of Italy from December 1963 to June 1968 and then from November 1974 to July 1976. Moro also served as Minister of Foreign Affairs from May 1969 to July 1972 and again from July 1973 to November 1974. During his ministry he implemented a pro-Arab policy. Moreover, he was appointed Minister of Justice and of Public Education during the 1950s. From March 1959 until January 1964, Moro served as secretary of the Christian Democracy. On 16 March 1978 he was kidnapped by the far-left terrorist group Red Brigades and killed after 55 days of captivity. He was one of Italy's longest-serving post-war Prime Ministers, leading the country for more than six years. An intellectual and a patient mediator, especially in the internal life of his own party, during his rule, Moro implemented a series of social and economic reforms which deeply modernized the country. Due to his accommodation with the Communist leader Enrico Berlinguer, known as the Historic Compromise, Moro is widely considered one of the most prominent fathers of the modern Italian centre-left and one of the greatest and most popular leaders in the history of the Italian Republic. (en)
Aldo Romeo Luigi Moro (Maglie, 23 settembre 1916 – Roma, 9 maggio 1978 è stato un politico, giurista e accademico italiano. Tra i fondatori della Democrazia Cristiana e suo rappresentante alla Costituente, ne divenne prima segretario (1959 e poi presidente (1976 e fu più volte ministro. Cinque volte Presidente del Consiglio dei ministri, guidò governi di centro-sinistra (1963-68 promuovendo nel periodo 1974-76 la cosiddetta strategia dell'attenzione verso il Partito Comunista italiano attraverso il cosiddetto compromesso storico.Fu rapito il 16 marzo 1978 e assassinato il 9 maggio successivo dalle Brigate Rosse. È uno dei quattro Presidenti del Consiglio dei Ministri della Repubblica italiana ad aver ricoperto questa carica per un periodo cumulativo maggiore di cinque anni. (it)
dbo:activeYearsEndDate
1957-05-15 (xsd:date)
1959-02-15 (xsd:date)
1968-06-24 (xsd:date)
1972-07-29 (xsd:date)
1974-11-23 (xsd:date)
1976-07-29 (xsd:date)
dbo:activeYearsStartDate
1955-07-06 (xsd:date)
1957-08-10 (xsd:date)
1963-12-04 (xsd:date)
1969-05-05 (xsd:date)
1973-07-07 (xsd:date)
1974-11-23 (xsd:date)
dbo:almaMater
dbr:University_of_Bari
dbo:birthDate
1916-09-16 (xsd:date)
dbo:birthPlace
dbr:Apulia
dbr:Kingdom_of_Italy
dbr:Maglie
dbo:deathDate
1978-05-09 (xsd:date)
dbo:deathPlace
dbr:Lazio
dbr:Rome
dbo:deputy
dbr:Ugo_La_Malfa
dbr:Pietro_Nenni
dbo:occupation
dbr:Professor
dbo:office
Minister of Foreign Affairs
Minister of Grace and Justice
Minister of Public Education
dbo:predecessorOffice
Prime Minister of Italy
dbo:party
dbr:Christian_Democracy_(Italy)

Figura 9 - Visualizzazione di una porzione della pagina Wikipedia relativa ad Aldo Moro attraverso il browser visuale di DBpedia.
Tratto da https://dbpedia.org/page/Aldo_Moro

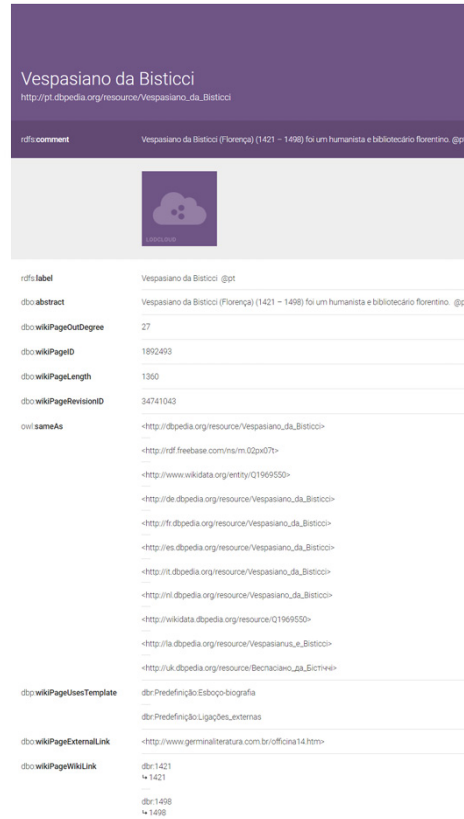

Figura 10 - Visualizzazione di una porzione della pagina DBpedia relativa a Vespasiano da Bisticci con LodView.
Tratto da http://pt.dbpedia.org/resource/Vespasiano_da_Bisticci

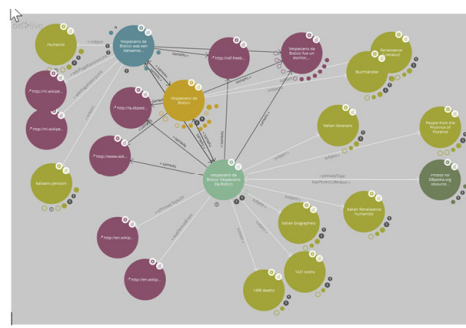

Figura 11 - Visualizzazione dell'istanza Vespasiano da Bisticci su DBpedia attraverso LodLive (grafo etichettato), http://lodlive.it

La navigazione a faccette è uno degli strumenti principe tanto per l'esplorazione quanto per la visualizzazione dei LOD proprio perché sfrutta le categorie descrittive della base di conoscenza come strumenti di browsing amichevole.

Ma dobbiamo aggiungere qualche ulteriore riflessione. Visualizzare i dati va oltre la loro navigazione testuale. Se con l'accesso per faccette si possono filtrare i risultati di una ricerca, affiancando la tradizionale interrogazione *google-like* con un'esplorazione per categorie, i dati della base di conoscenza possono rivelare nuovi risultati quando nuove strategie di visualizzazione vengano messe in atto.[272]

Tradurre stringhe di caratteri in elementi visuali ha una storia assai lunga che, dalle mappe geografiche ai grafici a torta, passando per l'albero enciclopedico delle conoscenze,[273] arriva alle recenti infografiche.[274] Le canoniche entità dei LOD (ovvero anche le già citate quattro w: who, when, where, what) diventano gli strumenti per una prima esplorazione visuale della conoscenza. Ecco allora che, in questi ultimi anni, si sono moltiplicate API, librerie e tool per fornire modalità diverse di visualizzazione.[275] Il tradizionale approccio attraverso indicizzazione per lista (ad es. persone, luoghi, lemmi, riferimenti bibliografici) è stato affiancato da nuovi strumenti di visualizzazione: torte o istogrammi per dati quantitativi rispetto a un numero finito di variabili, grafi per la gestione di network di relazioni, insiemi in caso di *cluster* o di interconnessioni fra *cluster*, fino alle più note e utilizzate linee temporali (*timeline*) per navigare gli eventi su un asse cronologico o ancora puntatori su una mappa per geolocalizzare entità rispetto al territorio.

Diciamo che la visualizzazione è prima di tutto uno strumento interpretativo. Ovvero certe metodologie sono più efficaci di altre per consentire all'utente di comprendere davvero il contenuto di una raccolta di risorse, in base alla tipologia di dati a disposizione. Non esiste in linea assoluta una visualizzazione giusta e una sbagliata, ma la scelta di quella visualizzazione capace di esprimere quello che i dati hanno da dirci sulla base delle loro caratteristiche e in relazione agli obiettivi della ricerca.[276]

Diremo anche che:

> Visualization techniques can be classified according to the **type of analytical task that the user is attempting to perform on the data**. Visualization techniques such as pie charts are appropriate for comparing the attributes or

> values of different variables within the dataset. If the user wants to analyse relationships and hierarchies then graphs and other related techniques can be used. The analysis of data in time or space can be supported with timelines and maps. A scatter plot can be used to analyse three-dimensional data. Higher dimensional data can be visualised using techniques such as radar charts.[277]

Visualizzare i dati significa quindi restituire forme di interpretazione così come emergono dalle triple (e dai modelli ontologici scelti). E gli strumenti della visualizzazione diventano strategie per la navigazione dei contenuti attraverso metafore visuali efficaci, come appunto usare il puntatore di un luogo per risalire agli item relativi a quel luogo o una linea del tempo con le date che scorrono su un asse orizzontale per richiamare gli eventi associati a quell'anno (Fig. 12). Questo ci fa capire che ci saranno allora potenzialmente tanti "chi", tanti "cosa", tanti "quando" e tanti "dove", quante le domande di ricerca che sarà possibile formulare sulle nozioni di persona, di oggetto/concetto/soggetto, spazialità e temporalità dei diversi dati e metadati necessari a identificare e qualificare le risorse.

Riprendendo un progetto già introdotto, ovvero l'ultima edizione semantica delle lettere di Vespasiano da Bisticci, in ottica di interazione con il lettore si sono scelte alcune visualizzazioni, utili a mostrare alcuni quesiti di ricerca in termini visuali:[278] 1) la visualizzazione geografica è servita per identificare sulla mappa i luoghi di spedizione e ricezione delle missive ("dove"); 2) la visualizzazione temporale su un asse orizzontale è servita per mettere in relazione le lettere con le date di ricezione e invio ("quando"); 2) le persone, sono state visualizzate attraverso un istogramma, per determinare il numero di volte che ogni corrispondente occorre, assieme a un diagramma ad arco per mostrare le relazioni fra le persone così come compaiono nelle lettere ("chi"); 3) il lessico tecnico della copia e della tradizione manoscritta, tipizzato per categorie, è stato mostrato con un grafico a torta per stabilire la percentuale di occorrenze della terminologia rispetto alle categorie ("cosa"). Le visualizzazioni così realizzate consentono di trarre una serie di considerazioni, come ad es. comprendere in che periodo le lettere sono più frequenti, o qual è il luogo da cui vengono spedite più lettere, chi è il corrispondente più assiduo o quali sono le persone che hanno più contatti reciproci, o ancora rispetto a quale categoria semantica il lessico è più

frequente.

E si entra così nel tema del *visual analytics*,[279] perché la visualizzazione è anche uno strumento di analisi. O meglio, gli esiti del tradizionale approccio condotto attraverso tecniche di *text retrieval* possono trasformarsi da testuali a visuali, soprattutto quando si aprano le prospettive all'analisi statistica dei risultati (ad es. con il linguaggio R[280]), verso il *machine learning*, alla scoperta di *pattern* nascosti all'interno del dataset.

Allora avremo a che fare con tool che sono suite di strumenti di analisi, che restituiscono un risultato visuale, perché è attraverso la visualizzazione, o meglio attraverso specifici modelli di visualizzazione, che si è in grado di comprendere fenomeni del testo.[281] Tool che mettono a disposizione API che possono essere integrate nel proprio ambiente per affiancare alla ricerca testuale un'esplorazione dei dati (*visual analytics* + *data exploration*) finalizzata a scoprire correlazioni nuove, svelare l'inaspettato, rivelare qualcosa di non conosciuto. Insomma, modificare le modalità di fruizione dei dati significa estendere la conoscenza, in particolare in un contesto di big data.

Ecco che al *close reading*, rappresentato dall'analisi puntuale del testo, ed espresso attraverso forme diverse di annotazione che documentano il piano qualitativo dell'analisi stessa, si sommano processi di *distant reading* del dato nei suoi contesti, in un approccio quantitativo che arricchisce l'esperienza di lettura, aggiungendo all'esperienza dell'ermeneuta la dimensione statistica e numerica dell'analisi.

Dalla *data visualization* passiamo quindi all'*information visualization* che è anche narrazione, creazione di percorsi, racconto di storie, come lo *storytelling* attraverso i LOD, che significa l'uso delle relazioni fra i dati come strumento per elaborare nuove strategie di lettura della conoscenza espressa dai beni culturali. Le nuove frontiere saranno allora mostre e tour virtuali, sistemi di realtà aumentata, esplorazioni in 3D, esperienze immersive e ricostruzioni virtuali, fino ai percorsi ipermediali aperti e collaborativi, capaci di integrare tecniche di gaming nel tradizionale approccio passivo ai dati del Web, anche a scopo educativo e didattico.[282]

Nuove strategie quindi per la disseminazione dei dati della ricerca per un'utenza profilata, con esigenze diversificate. Il punto di arrivo della progettualità nelle DH che si intende ora raccontare.

Chiudiamo allora con le parole di Johanna Druker che così sintetizza il concetto di visualizzazione come espressione dell'interpretazione condotta dall'umanista sui dati:

> rethinking **digital tools for visualization** on basic principles of the humanities. I take these principles to be, first, that the humanities are committed to the concept of **knowledge as interpretation**, and, second, that the apprehension of the phenomena of the physical, social, cultural world is through **constructed and constitutive acts**, not mechanistic or naturalistic realist representations of pre-existing or self-evident information.283

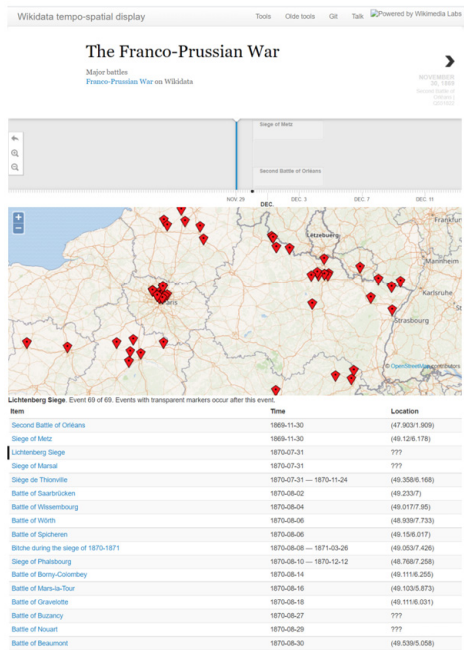

Figura 12 - Una semplice visualizzazione spazio-temporale.
Tratto da Wikidata Toolforge, https://tools.wmflabs.org/wikidata-todo/tempo_spatial_display.html

246 Eric Lease Morgan and LiAM, *Linked Archival Metadata: A Guidebook*, version 0.99, 2014, http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.
247 Claudio Gnoli, Vittorio Marino, Luca Rosati, *Organizzare la conoscenza:*

dalle biblioteche all'architettura dell'informazione per il Web, Pavia, Hops, 2006. Si veda poi il glossario di organizzazione della conoscenza, a cura di ISKO (International Society for Knowledge Organization) Italia, http://www.iskoi.org/doc/glossario.htm.

248 Francesca Tomasi, Knowledge Organization and Cultural Heritage, a.a. 2021/2022, https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2021/454462.

249 Digital Humanities and Digital Knowledge (DHDK), https://corsi.unibo.it/2cycle/DigitalHumanitiesKnowledge.

250 Birger Hjørland, *What is Knowledge Organization (KO)?*, "Knowledge Organization", 35 (2008), 2, p. 86-101, https://doi.org/10.5771/0943-7444-2008-2-3-86.

251 Su questi temi si veda in particolare Julia Flanders, Fotis Jannidis, *Knowledge Organization and Data Modeling in the Humanities*, [White Paper], 2015, https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/11127.

252 Russell Ackoff, *From Data to Wisdom*, cit.

253 Si veda in particolare Francesca Tomasi, *Edizioni o archivi digitali? Knowledge sites e apporti disciplinari*, in *Edizioni Critiche Digitali. Digital Critical Editions. Edizioni a confronto/comparing editions*, a cura di Paola Italia, Claudia Bonsi, Roma, Sapienza Università Editrice, 2016, p. 129-136.

254 Si veda Francesca Tomasi, *Digital humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM*, "Aibstudi", 60 (2020), 2, p. 1-15, https://doi.org/10.2426/aibstudi-12068.

255 Cristina Pattuelli, *Mapping People-centered Properties for Linked Open Data*, "Knowledge Organization", 38 (2011), 4, p. 352-359. Grassetto nostro.

256 Si veda, ad es., il survey sulla qualità dei LOD in Gustavo Candela et al., *Evaluating the Quality of Linked Open Data in Digital Libraries*, "Journal of Information Science" (August 2020), p. 1-23, https://doi.org/10.1177/0165551520930951; ma non mancano posizioni non così favorevoli al fenomeno LOD così come fino ad oggi è stato condotto, per cui si veda, ad es., Kyle Banerjee, *The Linked Data Myth*, "Library Journal", News (August 13, 2020), https://bit.ly/2VIXPnQ.

257 Un riferimento utile è Gail Hodge, *Systems of knowledge organization for digital libraries. Beyond traditional authority files*, Washington, DC, The Council on Library and Information Resources, 2020, http://www.clir.org/pubs/reports/pub91/contents.html. Si veda in particolare il cap. 2. *Linking Digital Library Resources to Related Resources*.

258 Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, 2012, https://public.ccsds.org/pubs/650x0m2.pdf.

259 Si veda, in particolare, l'attività del W3C Working Group, *Best Practices for Publishing Linked Data*, 2014, https://www.w3.org/TR/ld-bp.

260 CIDOC CRM, SPAR ontologies e HiCO, assieme ai vocabolari controllati e agli authotity files di AAT per la terminologia storico/artistica; VIAF e ULAN per i nomi di persona; Dbpedia e Wikidata per diversi parametri di allineamento e GeoNames per i luoghi geografici.
261 L'ultima normativa per la descrizione delle Fotografie (F) è alla versione 4.0, 2016, http://www.iccd.beniculturali.it/getFile.php?id=8144.
262 L'ultima normativa per descrizione di Opere e oggetti d'arte (OA) è alla versione 3.0, 2018, http://www.iccd.beniculturali.it/getFile.php?id=7508
263 F Entry Ontology, https://w3id.org/people/essepuntato/fentry.
264 OA Entry Ontology, http://purl.org/emmedi/oaentry.
265 Si veda Agenzia per l'Italia Digitale (AGID), *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, cit., p. 23 del formato .pdf.
266 Marilena Daquino, Francesca Giovannetti, Francesca Tomasi, *Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini*, "Umanistica Digitale", 3 (2019), 7, https://doi.org/10.6092/issn.2532-8816/9091.
267 Aldo Moro, *Edizione Nazionale delle Opere di Aldo Moro*, Bologna, Università di Bologna, 2021, https://doi.org/10.6092/unibo/aldomoro. L'edizione digitale si può navigare nell'ambiente https://aldomorodigitale.unibo.it. I principi che hanno governato la modellazione dei dati si possono leggere nel report tecnico a cura di Sebastian Barzaghi, *La modellazione dei dati nell'Edizione Nazionale delle Opere di Aldo Moro*, Zenodo, 2021, https://doi.org/10.5281/zenodo.5144961, p. 13-17.
268 Flusso di lavoro, criteri editoriali, modellazione dei dati, elaborazione dei documenti, sito dell'edizione e risultati raggiunti si possono leggere alla URL https://aldomorodigitale.unibo.it/about.
269 Il flusso di lavoro dell'edizione delle opere di Aldo Moro si può leggere alla pagina https://aldomorodigitale.unibo.it/about/docs/workflow#survey-transcription-section, che riporta un grafico del workflow (a cura di Sebastian Barzaghi).
270 Sul rapporto fra sistemi di organizzazione della conoscenza e CC, si veda Carlo Bianchini, *Organizzare la conoscenza con la sequenza di filiazione della Classificazione Colon di S.R. Ranganathan*, "JLIS.it", 2 (2011), 2, p. 1-2, https://doi.org/10.4403/jlis.it-4710.
271 Si veda, ad es., Francesca Tomasi, Fabio Ciotti, Marilena Daquino, Maurizio Lana, *Esplorare semanticamente collezioni culturali: uno studio di fattibilità*, "AIDAinformazioni", 3-4 (2015), p. 125-143, https://doi.org/10.4399/97888548899278.
272 Si veda in particolare il recente volume Laura Po, Nikos Bikakis, Federico Desimoni, George Papastefanatos, *Linked Data Visualization: Techniques, Tools, and Big Data*, "Synthesis Lectures on the Semantic Web: Theory and Technology", 10 (2020), 1, https://doi.org/10.2200/

S00967ED1V01Y201911WBE019.

273 Uno dei più noti esempi *ante litteram* di *data visualization* è l'albero enciclopedico delle conoscenze, https://en.wikipedia.org/wiki/Figurative_system_of_human_knowledge.

274 Celeberrima la "geniale rappresentazione della campagna di Russia di Napoleone realizzata da Charles Joseph Minard, considerata da Edward Tufte, uno dei massimi esperti contemporanei in materia, la migliore visualizzazione statistica di sempre", https://www.visualitica.it/storia-delle-infografiche-charles-joseph-minard.

275 Solo per nominarne qualcuno: applicazioni Web che facilitano la creazione di strumenti visuali, come Histropedia per costruire linee temporali (http://histropedia.com/timeline-everything.html) o i tool di Knightlab per la realizzazione di *storytelling* (https://knightlab.northwestern.edu/projects/); set di strumenti di visualizzazione come RAWGraphs (https://rawgraphs.io) o Gephi per le reti sociali (https://gephi.org); librerie Javascript come Leaflet (https://leafletjs.com), D3 (https://d3js.org), DIMPLE (http://dimplejs.org) o C3 (https://c3js.org).

276 Una classificazione degli strumenti di visualizzazione, *clusterizzabili* a seconda della tipologia di attività che si intende svolgere (confronto, relazioni, distribuzione o composizione) e a seconda del tipo di dati che si hanno a disposizione e il conseguente numero di variabili, si può leggere in *Data Analysis And Visualization – Are You Doing it Right?*, https://eazybi.com/blog/data-analysis-and-visualization.

277 *Linked Data Visualization*, in Enrico Fagnoni et al., *How to use Linked Data*, 2015, https://en.wikitolearn.org/Course:How_to_use_Linked_Data/Interaction_with_Linked_Data/Linked_Data_Visualization. Grassetto nostro.

278 Si veda la descrizione analitica delle scelte di visualizzazione, nella pagina dedicata alla documentazione del sito, a cura di Sebastian Barzaghi, https://dharc-org.github.io/vespasiano-da-bisticci-letters-de/documentation/visualization.html.

279 Si tratta di passare dalla visualizzazione ai sistemi per l'analisi dei risultati. Uno strumento molto usato è Tableau, https://www.tableau.com.

280 Un interessante tool francese è Iramuteq, "Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. Un logiciel libre construit avec des logiciels libres", http://www.iramuteq.org.

281 Fra i vari strumenti, vale senz'altro la pena menzionare Voyant Tools "a web-based text reading and analysis environment. It is a scholarly project that is designed to facilitate reading and interpretive practices for digital humanities students and scholars as well as for the general public", https://voyant-tools.org. Ma si veda il progetto Alcide (Analysis of Language and Content In a Digital Environment), piattaforma realizzata della Fondazione Bruno Kessler (FBK), per l'analisi dei testi e utilizzato per l'edizione Alcide De Gasperi, *Scritti e discorsi politici*, I-IV, sotto la direzione scientifica di Paolo Pombeni, Bologna, il Mulino, 2006-2009, https://

alcidedigitale.fbk.eu.

282 Su questi temi, che non approfondiamo oltre, si può senz'altro vedere l'attività svolta dal centro Framelab dell'Università di Bologna, con i progetti realizzati, https://framelab.unibo.it/index.php/progetti. Interessante anche la sezione del blog, in particolare sullo *storytelling*, https://frameblog.unibo.it/index.php/category/storytelling e sui *serious games*.

283 Johanna Drucker, *Humanities Approaches to Graphical Display*, "Digital Humanities Quarterly", 5 (2011), 1, http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html. Grassetto nostro.

6. PROGETTARE RISORSE DIGITALI NEL CULTURAL HERITAGE

6.1 DH e lavoro a progetto

Una delle specificità delle DH, come si sarà ormai compreso dai capitoli precedenti, è sicuramente il fatto che molte delle attività portate avanti dai suoi ricercatori sono finalizzate alla realizzazione di un progetto. Che sia l'edizione scientifica digitale (diplomatica, interpretativa o critica, per usare una classificazione canonica e condivisa) di un testo (o della sua tradizione), una collezione di documenti storici di natura archivistica, una raccolta di artefatti culturali di origine museale, un insieme di oggetti, anche eterogenei per supporto, formato e tipo, o ancora un tool necessario ad assolvere un bisogno pratico, ma anche un modello concettuale da tradurre in un'ontologia formale, le DH trovano nella progettualità la naturale modalità di impiegare metodologie computazionali per la valorizzazione dei dati del vasto dominio delle scienze umane.[284] Diremo anche che il *digital humanist* è colui o colei che sa riconoscere i bisogni – molteplici, complessi e sfaccettati – della ricerca umanistica e che al contempo è consapevole delle metodologie computazionali più adeguate per assolverli. E realizzare un progetto.

Non è un caso che il corso di laurea DHDK, attivato nel 2017 presso l'Università di Bologna, abbia individuato nel lavoro a progetto, anche collaborativo, la modalità di valutazione dell'apprendimento dei corsi attivati in questa laurea magistrale internazionale.[285] Processo ad apprendimento incrementale: nuove conoscenze e competenze, acquisite progressivamente nel percorso formativo, possono arricchire il risultato di un progetto, allo scopo di arrivare alla tesi finale con tutti gli strumenti necessari a

realizzare un lavoro compiuto. Le nozioni di *computational thinking , data science* e i linguaggi di programmazione, assieme a quelle relative alle basi di dati, le tecnologie Web, i sistemi di rappresentazione e organizzazione della conoscenza in archivi, biblioteche e musei, il *digital heritage* e la museologia, le edizioni digitali assieme ai sistemi di *retrieval, analysis and mining* del testo si arricchiscono, dopo il primo anno, con le competenze su *business strategy* e autoimprenditorialità, diritto – e quindi problemi relativi all'accesso aperto e al copyright – e comunicazione, con un'enfasi in particolare su pratiche di *storytelling* digitale e di *information visualization* assieme alla *(social) network analysis*. Per arrivare a una conoscenza globale dei temi delle DH che possa mettere in condizione studenti e studentesse di realizzare un progetto di settore adeguato sul piano teorico, metodologico e tecnico.[286]

Tradizionalmente sappiamo che esistono alcune metodologie che sono condivise dalle diverse discipline umanistiche. I linguaggi di markup, le basi di dati, i linguaggi del Web semantico, i problemi di modellazione o classificazione, l'ipertestualità e la multimedialità sono alcune delle metodologie che attraversano trasversalmente ogni azione riconducibile alle scienze umane – in particolare trascrizione, o in generale acquisizione e creazione di dati, attività di analisi, interpretazione, descrizione, arricchimento, divulgazione e conservazione. Ma cosa hanno in comune i progetti di DH?

Proviamo dunque ad andare per gradi. È con l'analisi di cosa sia un progetto nelle DH che Johanna Drucker apre il suo volume DH101,[287] proprio per iniziare ad affrontare il tema delle DH, e i concetti che sono correlati a questo dominio, partendo dall'output, ovvero anche dal progetto finito, con una tecnica di decostruzione delle sue componenti. Questo significa che, prima di parlare di conoscenze e competenze, questo volume analizza il risultato finale:

> At their simplest, digital projects can consist of a set files (**assets**) stored in an information architecture such as a database or file system (**structure**) where they can be accessed (**services**) and called by a browser (**use/display**).

Sintetizzando, continua la Druker, un progetto di DH si può schematizzare nei seguenti elementi base:

- a repository of files or digital **assets**,
- some kind of information architecture or **structure**,
- a suite of **services**,

- and a **display for user experience**.

Ne deduciamo che ci potranno bastare come base di partenza: una collezione di file, un luogo per conservare questi file, degli strumenti di accesso e la possibilità di poter fruire di questi file da parte di un utente finale. Nulla di più semplice. Potremmo aggiungere qualche elemento in più, giusto per iniziare a dare consistenza all'argomentazione. In prima battuta ogni progetto parte da un'idea. Scopo, obiettivo e finalità del progetto devono essere ben chiari. Dopodiché certamente la costruzione di una collezione di risorse (diciamo testi, immagini, volendo file audio o video), già digitali o da digitalizzare, costituirà il *repository*, ovvero il luogo di deposito, esito di una selezione degli oggetti necessari ad assolvere il bisogno informativo emerso dell'idea progettuale. Oggetti che andranno adeguatamente descritti, ovvero metadatati. Aggiungeremo che questa collezione (dati e metadati) andrà caricata da qualche parte, ovvero dovrà essere consultabile. Se il punto di partenza è il disco rigido del computer, dare struttura, come abbiamo visto, significa scegliere il modello di dati più adeguato a rappresentare gli oggetti della nostra collezione, raccolti nel *repository*. Ed ecco la fase più delicata del progetto, perché da questa scelta ne conseguono tanto i servizi che è possibile erogare all'utente finale, quanto le modalità di accesso alla collezione. E le modalità di accesso, o meglio di fruizione, sono un tema particolarmente delicato. Perché sappiamo che il luogo più naturale di pubblicazione di un progetto di DH è il Web, ma pubblicare sul Web richiede, oltre alla scrittura delle pagine che popolano il sito, di dialogare con il *repository*, per fornire all'utente finale dati accessibili in dimensione *user-oriented*. Soprattutto se l'idea del progetto è rendere accessibili i file attraverso un sito Web. Non solo quindi l'interfaccia richiederà attenzione, e cura particolare, ai principi di architettura dell'informazione, *Web design* e layout grafico, *labeling*, usabilità, pratiche di browsing innovative (spesso dipendenti dal tipo di classificazione dei dati esito della metadatazione), ma anche al dialogo con il dataset, diciamo cioè la collezione dei file dotata di una qualche struttura, vieppiù se ci troviamo a utilizzare gli strumenti del Web semantico.

Un processo dunque che, partendo da una serie di step apparentemente semplici e lineari, nasconde una serie di complessità, in ogni momento della progettazione, perché richiede

l'assunzione di un punto di vista che, a sua volta, stabilisce metodi e strumenti più adeguati per affrontare ogni fase del *workflow*.

Dato questo percorso astratto, quando ci si confronti con la realtà dei progetti realizzati nel settore delle DH si rileva un ulteriore complessità: come chiamare i progetti che condividono questo *workflow*? Non è senza dubbio semplice classificare un prodotto di DH. Se andiamo a leggere le descrizioni dei progetti presenti in EADH, ci scontriamo con una serie di definizioni: se la terminologia più ricorrente è quella di archivio ed edizione, collezione e catalogo, non mancano i termini inventario, biblioteca, repertorio, dizionario, enciclopedia, raccolta, a cui aggiungiamo le nozioni di tool, infrastruttura, ambiente. Ovviamente digitale.

Al di là delle motivazioni per queste differenze terminologiche, che danno comunque il tenore di un panorama ricco ed eterogeno, quello che si evince dalla consultazione dei progetti di DH è la difficoltà a tipizzare i diversi prodotti. Proprio perché in un contesto fluido come quello del Web le tipologie di risultati si scontrano con il tentativo di stabilire tassonomie rigorose, da usare come strumento di classificazione e valutazione. La comunità sta dimostrando un interesse importante in questo ambito, testimoniato anche da un significativo sforzo di portare avanti iniziative; particolarmente utile, ad es., la ricerca condotta in alcuni lavori, finalizzati proprio alla classificazione di attività, oggetti e tecniche/tecnologie nel settore delle DH (concetti sui quali torneremo nell'ultima sezione di questo capitolo).[288]

Ma data la complessità a dare la giusta tipologia ad un prodotto, è senz'altro utile analizzarne le componenti che, andando oltre l'attribuzione terminologica, ci permettano di coglierne l'essenza; e farlo attraverso un processo di *reverse engineering* che ci permetta di decomporre un progetto di DH.[289] Ovvero possiamo cercare di rispondere alla domanda: come è stato fatto questo progetto digitale che sto navigando? Perché tipicamente la forma naturale del nostro output, l'abbiamo detto, è il Web. E come operare in questa direzione se non cercando di capire quali siano i parametri da rispettare per fare un progetto di DH degno di questo nome?

Un primo suggerimento ci viene da RIDE,[290] che è sì una rivista che raccoglie recensioni di prodotti digitali eterogeni della ricerca (edizioni, collezioni di testi e tool), ma è anche classificabile come un repertorio di linee guida, le quali, presentandosi come gli strumenti necessari ad effettuare il lavoro di revisione, stabiliscono

le buone pratiche per un progetto di DH[291]. Senza la pretesa di un'analisi puntuale delle linee guida, certamente colpiscono alcuni elementi che devono guidare il recensore nella valutazione del progetto: 1) la possibilità di riconoscere un'identificazione bibliografica precisa, assieme alla capacità di qualificare chiaramente ruoli e funzioni di tutte le entità (persone, enti e istituzioni) coinvolte nel progetto; 2) l'attenzione per la chiarezza nel determinare, e quindi comprendere, oggetto e contenuto, assieme alla riconoscibilità dei metodi che il progetto dichiara di utilizzare e degli obiettivi che intende perseguire; 3) l'importanza degli aspetti di pubblicazione e presentazione del progetto, compresi quelli visuali.

La lezione che deriviamo da RIDE, fra le altre, è che un'azione fondamentale in ogni progetto di DH è affiancare ai contenuti di progetto una documentazione, anche a livello tecnico, esaustiva. Documentare un progetto significa tanto descrivere ogni step del *workflow* che si è seguito per arrivare dal processo al prodotto, quanto consentire all'utente interessato di replicare lo stesso prodotto (e processo) con dati diversi. La documentazione per un progetto di DH è come la nota al testo per un'edizione: deve mettere in condizione il lettore/fruitore di risalire al ragionamento che ha guidato l'editore critico alla ricostruzione del testo e della sua tradizione, o degli stati della sua evoluzione. I dettagli di un lavoro digitale sono dunque lo strumento di dialogo con la comunità, rappresentano la volontà di raccontare le scelte del progettista e di condividere un approccio critico a dati e risorse, ma anche a strumenti e metodi. Modello del testo e modello dei dati, tool e tecnologie, sistemi di visualizzazione e componenti di interfaccia sono le basi su cui si fonda la documentazione, e sono i concetti che guidano il lettore alla comprensione del prodotto, attraverso la disamina delle fasi del processo.

La descrizione delle componenti che sono state utilizzate durante il processo rende infatti il prodotto finale riproducibile. E la replicabilità è un tema fondamentale nel Web semantico. Tanta letteratura è dedicata oggi al tema dei dati FAIR (*Findability, Accessibility, Interoperability, Reuse*),[292] la strategia finalizzata alla diffusione degli open data. Reperibilità, accessibilità, interoperabilità e riuso sono cioè i punti cardine che devono guidare la produzione di dati aperti, ovvero: 1) la loro identificazione univoca affinché siano trovabili; 2) l'uso di

protocolli aperti per la massima accessibilità; 3) l'impiego di metadati standard e vocabolari controllati per l'interoperabilità; 4) la pubblicazione dei dati con licenze aperte e dichiarazione di provenienza a garanzia di corretto riutilizzo. E se tanto si parla di dati (e di conseguenza di metadati) FAIR, non va dimenticata la cura da porre nell'adesione ai principi FAIR anche nei processi che hanno condotto alla realizzazione del prodotto. E perché dati, metadati e processo, o anche il *workflow*, siano FAIR, anche il vocabolario, ovvero l'ontologia, dovrà seguire le stesse regole,[293] così come l'infrastruttura che ospita dati, metadati e vocabolari e che sta alla base del processo di creazione del dataset. E tutto il percorso della "FAIRification" dovrà essere adeguatamente documentato. Strumenti come DCAT,[294] un vocabolario RDF per descrivere raccolte/collezioni/cataloghi di dati sul Web, stanno contribuendo, ad es., a far condividere all'interno della comunità l'importanza di descrivere dataset, attraverso opportuni metadati, ma anche di documentare servizi erogati dagli ambienti o messi a disposizione dalle applicazioni, al fine di garantire massimo interscambio.

La documentazione dunque è alla base del riuso. Ed è innegabile che grande sforzo sia oggi rivolto alla creazione di ambienti, tool e infrastrutture che consentano il riuso tanto dei dati quanto delle applicazioni, dei vocabolari e di tutte le componenti che sovrintendono alla creazione di un progetto digitale. E per proseguire con il nostro approccio alla *reverse engineering* dobbiamo riflettere sulle buone pratiche, o anche sulle linee guida che possono fornire le specifiche per un percorso nelle DH, ma al contempo che sono capaci di consentirci di ragionare in termini di decostruzione di progetti – e dunque dati, processi e servizi – esistenti.

6.2 Buone pratiche: un *workflow* di progetto

Se tipicamente un progetto di DH trova nel Web il suo naturale luogo di pubblicazione, è di nuovo il Web il luogo eletto ad accogliere anche la documentazione descrittiva del progetto realizzato. È allora buona norma che ogni progetto sia accompagnato da un *Web project plan*, ovvero dalla serie delle informazioni necessarie all'utente finale per comprendere ragioni,

scopi, obiettivi, strumenti e metodi del progetto che sta consultando.

Come abbiamo detto, ogni progetto di DH deve ovviamente partire da un'idea di ricerca. Ovvero anche da un obiettivo culturale da raggiungere e una finalità da perseguire. Ogni progetto avrà una tipologia di riferimento, che determinerà tanto il/i formato/i dei dati, quanto le strutture più adeguate a rappresentarli. L'ambiente di fruizione ricoprirà un ruolo fondamentale per fornire servizi di accesso e visualizzazione dei dati. Dovranno essere stabiliti sistemi di preservazione a lungo termine e di archiviazione. Riconoscimento della paternità delle azioni e in generale ruoli delle persone coinvolte nel progetto (*credits*), indicazioni su come citare il lavoro ("how to cite"), coordinate temporali e spaziali (luogo e data di pubblicazione), messa a disposizione della comunità del codice (strumenti di *sharing*), meccanismi di identificazione univoca (come URI o DOI), licenze e criteri per la fruizione (CC-BY) sono i principali elementi che nessun progetto di DH può dimenticare come corredo paratestuale necessario a garantire qualità e affidabilità dell'informazione.

Proviamo allora ad elaborare uno scheletro di riferimento, ovvero individuiamo i punti che ogni documento che accompagni un progetto di DH non può mancare di menzionare, e aggiungiamo gli interrogativi che è necessario porsi per compilare ogni voce del documento. Questo elenco di punti di attenzione dovrà essere seguito quando si intenda realizzare un progetto, ma può al contempo essere pensato come l'insieme degli elementi di controllo per la valutazione critica di una risorsa esistente.

1. Brief. L'analisi

- Obiettivi, finalità, scopo. Che tipo di intento vuole raggiungere la risorsa digitale? Quali sono gli obiettivi del progetto, quali gli scopi della ricerca, quali le domande alle quali il progetto vuole rispondere?
- Utente tipo. Qual è l'utente tipo? Possono esistere tipologie di utenti diversi? Come cambia l'accesso alle risorse in base alla profilazione?
- Modalità di accesso. Per che tipo di piattaforma è pensata la risorsa? Desktop, tablet, mobile? Come cambierà l'accesso ai contenuti in base al *device*?

- Contenuti. Di cosa parla la risorsa? Dove è possibile reperire i contenuti? Le fonti di riferimento sono identificabili? I contenuti sono già in formato digitale o devono essere acquisiti? E in questo ultimo caso sono chiari i parametri (ad es. licenze e formati)?

2. Benchmark. L'esistente

- Esistente e *competitors*. Esiste già qualcosa che parla dello stesso argomento? In cosa si distingue il progetto che si intende realizzare rispetto a quanto già reperibile in rete? Quali elementi positivi e quali negativi hanno le risorse concorrenti? Cosa è da salvare e quindi cosa è da riusare rispetto all'esistente?

3. I dati. Costruire la raccolta

- Tipologia. Con quale tipologia di dati si ha a che fare? Testi, immagini, audio, video? Documenti d'archivio, libri, descrizioni catalografiche, artefatti, riviste? Che tipo di relazioni ci sono fra i dati coinvolti nel progetto? In quali formati è possibile rappresentare i dati?
- Standard di riferimento. Esistono standard di metadatazione già previsti per le tipologie di dati che si intendono trattare? Se esistono, come è possibile riusarli? Se non esistono come va creato un set di descrittori ad hoc per le risorse? Esistono vocabolari controllati già disponibili per il tipo di dati con cui si ha a che fare?
- Categorie. È possibile categorizzare le risorse sulla base di tipologie preordinate? Se si dispone di risorse eterogenee per supporto e tipo è possibile individuare delle classi comuni per descrivere gli item della raccolta (ad es. Dublin Core)?

4. Struttura e modello dei dati

- Markup. Sulla base della tipologia di dati ha senso usare sistemi di markup? E se sì, va bene un markup *embedded* o è necessario ricorrere a sistemi *stand-off*? E a quale grammatica è possibile riferirsi? Va bene uno Schema esistente?
- DB. Sulla base della categorizzazione dei dati è bene usare un database? Per quale parte dei dati? È sufficiente un database

relazionale? Avrebbe senso ricorrere ad un database a grafo?
- Tecnologie Semantiche. È possibile valorizzare il database o la collezione di file marcati attraverso l'impiego di tecnologie semantiche? Avrebbe senso creare un dataset in LOD? Quali ontologie si potrebbero usare in questo caso?

5. Mappa e schema

- Mappa dei concetti. Sulla base dell'analisi dei contenuti è necessario produrre un sistema di interconnessione fra i concetti che stanno alla base del contenuto delle risorse. Può un modello E/R essere utile per schematizzare entità, attributi e valori?
- Schema delle dipendenze. Una volta concettualizzata l'idea e realizzati i dati e i metadati, come è possibile creare le pagine da rendere disponibili per la fruizione? Ovvero come è possibile realizzare il sito Web, a partire dalle risorse e dai dati, stabilendo le gerarchie delle pagine in una struttura ad albero?

4. L'interfaccia. Il layout Web

- Architettura. Che caratteristiche deve avere l'interfaccia del sito? Testata, navbar, sezioni, navigazione contestuale, navigazione primaria, footer. Si ha chiaro come va realizzata, dal punto di vista dell'architettura logica di interfaccia, l'home e di come andranno realizzate le altre pagine?
- *Wireframes*. Una volta stabilita l'architettura è possibile passare al disegno dei *wireframes*. Si ha chiaro di quanti *wireframes* si ha bisogno sulla base di quanti *templates* di pagina si immagina di realizzare?

5. Usabilità dell'interfaccia

- Efficacia e comunicabilità. Si ha chiaro cosa significa usabilità di un'interfaccia? Sia dal punto di vista dell'organizzazione dei contenuti che da quello dell'efficacia dei processi comunicativi in ambiente Web? Si conoscono le norme per un adeguato *user interface design*?
- Browsing. Come si rende navigabile il sito? Si conoscono le regole per la creazione di canali e i sotto-canali di accesso

(navigazione primaria, secondaria, contestuale, metanavigazione, briciole di pane)? Il *labeling* ipotizzato è efficace?
- Tipografia per il Web. Si conoscono le regole per la scrittura per il Web? E gli accorgimenti tipografici? Ruolo dei grassetti, gerarchie dei titoli, elenchi, liste, tabelle?
- Web grafica. Si conoscono i principi della grafica per il Web? Icone, colori, contrasti, font, spaziature, posizionamento?

6. Servizi. *User centered design*

- Strumenti di navigazione. Come si naviga il sito oltre ai canali primari? Faccette, strumenti di visualizzazione come mappe, linee del tempo, ricerca per parole possono essere utili? Esistono già delle componenti pubbliche che è possibile integrare?
- Strumenti di interazione. Quale interazione con le pagine e gli elementi in pagina? *Card, accordion, button, overlay, magnifier, carousel, hover, tooltip* ecc. possono essere usati? Quali sono i *widgets* più appropriati per l'interazione con la risorsa e le diverse tipologie di messaggi che si vogliono comunicare?
- Strumenti di visualizzazione. Sistemi di visualizzazione basati sull'analisi dei dati possono essere realizzati? Grafi, torte, istogrammi, avrebbero senso? Se sì, per quali tipologie di domande di ricerca da sintetizzare graficamente?
- Strumenti di esplorazione. È possibile immaginare itinerari e percorsi guidati? Elaborare mostre virtuali tematiche? Costruire uno *storytelling* profilato? Usare tecniche di *gamification*?
- Riuso API. Esistono tools che è possibile riusare nel progetto? Qualcun'altro ha già pensato a sviluppare applicazioni che potrebbero servire per il progetto che si intende realizzare (ad es. Voyant Tools per fare *text analysis*)?

7. Bibliografia

- Fonti. L'utente ha modo di verificare tutti i riferimenti bibliografici e sitografici necessari a comprendere da dove vengono le risorse nel progetto?
- Strumenti. L'utente ha tutti i riferimenti alle buone pratiche per un progetto digitale: URI (ad es. Figshare), DOI (ad es.

Zenodo), stringa bibliografica (possibilmente in diversi modelli o formati citazionali), dichiarazione di paternità (autori, ruoli e istituzioni), datazione (e date delle versioni o release) e localizzazione (se rilevante), licenze (per testi, dati e immagini), *repository* (ad es. Github), tools e librerie usate?

Giusto per fare un esempio in conclusione di questa riflessione, il già citato lavoro per la creazione dell'edizione delle opere di Aldo Moro, è partito proprio da un'analisi dell'esistente in termini di edizioni digitali, allo scopo di verificare le buone pratiche in uso dalle comunità e impiegare queste buone pratiche per stabilire i requisiti di qualità di un'edizione.

In particolare, riprendendo la documentazione prodotta per il progetto, le edizioni digitali che rispondono in modo più compiuto ai requisiti di qualità presi in esame presentano le seguenti caratteristiche:

- si rivolgono a un pubblico misto di specialisti e utenti generici;
- hanno una documentazione ricca e accurata;
- hanno una metadatazione completa;
- sono univocamente identificate e citabili, così come le parti che le compongono;
- sono arricchite da materiali aggiuntivi che ne contestualizzano il contenuto;
- si basano su un modello di dati orientato all'interoperabilità e all'interconnessione tra le risorse proprie e quelle già esistenti sul Web;
- usano strumenti di visualizzazione e *storytelling* per veicolare in maniera intuitiva le informazioni in esse contenute;
- sono basate su un sistema di organizzazione dell'informazione solido, indicizzato e facilmente navigabile;
- sono dotate di meccanismi di ri-utilizzo ed integrazione dei dati da parte di altri agenti (umani e macchine);
- possono essere integrate nell'ecosistema dei social media;
- hanno contenuti e dati scaricabili in molteplici formati differenti;
- si basano su codici e software con licenza Open Source;

- infine, hanno contenuti aperti e accessibili distribuiti con licenze che ne consentano il riutilizzo.[295]

Questi principi, che hanno governato l'analisi dell'esistente a scopo di *benchmark*, sono anche diventati gli elementi base da cui avviare il percorso di creazione dell'edizione digitale.

6.3 Il ciclo vitale degli oggetti culturali

Abbiamo oramai chiaro che quando si abbia a che fare con la progettualità nelle DH è inevitabile si ragioni su quelli che sono i momenti costituivi del ciclo vitale degli oggetti digitali di natura culturale (Fig. 13).

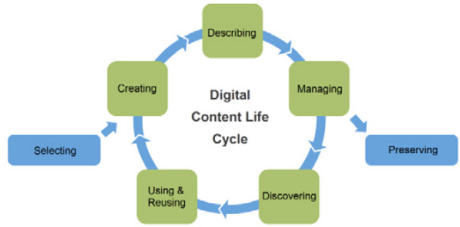

Figura 13 - Un possibile modello di *Digital Content Lifecycle*.
Tratto da https://localcontexts.org/traditional-digital-content-lifecycle

Il ruolo delle DH in questo contesto è di provvedere ai bisogni della ricerca, di fornire gli strumenti necessari a trasformare i dati in risorse scientificamente valide, attendibili, riusabili, capaci di documentare il tema della complessità, dell'interpretazione, del molteplice.

Abbiamo ben chiaro che dobbiamo affrontare il ciclo vitale di oggetti che sono eterogenei per supporto, formato e provenienza (archivi, biblioteche e musei). E che possono essere strutturati, semi-strutturati e non strutturati. Come sono eterogenei i dati, perché lo sono gli oggetti, lo sono le attività che possono essere condotte su tali oggetti e le tecniche che in ogni fase del ciclo vitale possono essere utilizzate (per cui si veda Tabella 2).

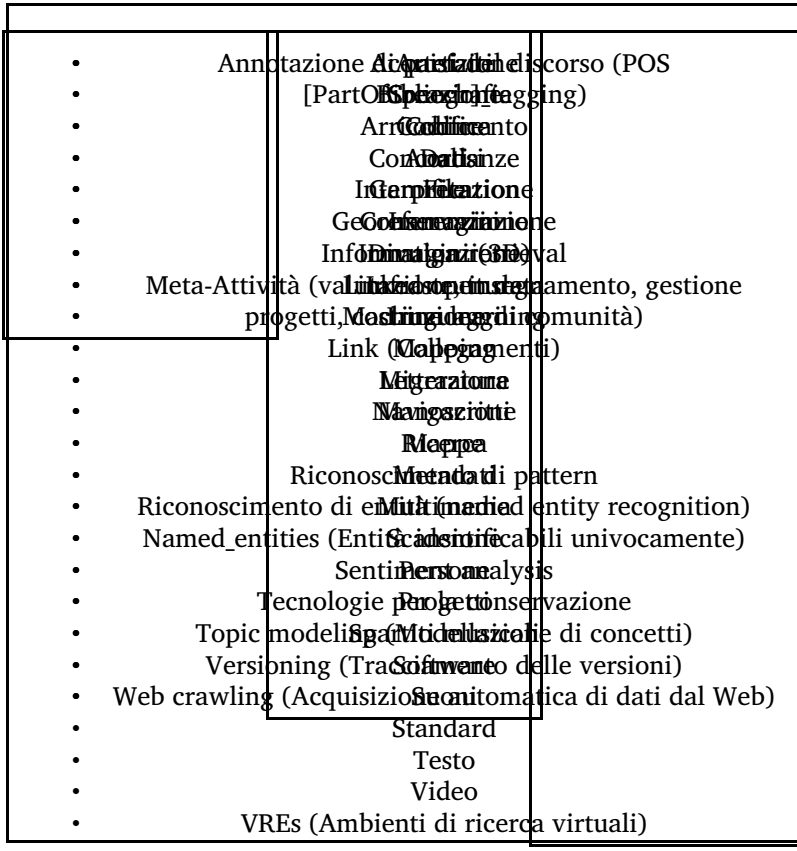

Tabella 2 - TaDiRAH - Taxonomy of Digital Research Activities in the Humanities, http://tadirah.dariah.eu/vocab. Selezione e traduzione nostra

Sicuramente, il concetto che sta a fondamento della nozione di ciclo vitale è quello di processo, che stabilisce le azioni, con la consapevolezza che ogni step di questo processo ha problemi di scelta e selezione in termini di: formati, strumenti, modelli, metodi e tools. Ma anche con la presa di coscienza che ogni momento del ciclo di vita potrebbe richiedere competenze e professionalità diverse.

Partiamo allora dalle azioni del ciclo vitale, correlandole a quelle stabilite dal modello TaDiRAH:

- Produrre (acquisizione e creazione): se ogni progetto parte dalla selezione degli oggetti che si intendono manipolare ai fini del processo, la prima azione è quella di acquisire

digitalmente e/o trascrivere tali oggetti, con la consapevolezza della necessità di far fronte alla questione dei diversi formati (sulla base della tipologia di fonte, ma anche in base allo scopo dell'azione di acquisizione, ad es. preservazione o disseminazione), tipologie di licenze (sia delle fonti da acquisire, che di rilascio dei dati), qualità dell'azione di acquisizione (ad es. diverse risoluzioni e apparecchi di scansione in caso di immagini o grado di complessità dell'operazione di annotazione in caso di formato testo). Questa prima fase ci consente la creazione dei dati.

- Descrivere (arricchimento): i metadati. Abbiamo già parlato del ruolo della descrizione e del momento fondamentale della modellazione. Non esiste dato privo di metadato. I LOD ci insegnano a produrre una nuova tipologia di dato che porta con sè la serie delle relazioni necessarie a descriverlo in modo analitico. È evidente che i vocabolari utilizzabili in questa fase sono numerosi e il modello di riferimento cambia, come cambia il punto di vista di chi modella e realizza poi il dataset. Abbiamo già detto che se arricchimento in contesto LOD significa allineamento a record di autorità, esso va pensato anche nei termini della costruzione di nuove relazioni fra entità. Lo sforzo andrà dunque condotto nel cercare di costruire nuove connessioni fra dati e di far dialogare la conoscenza in un circuito integrato.
- Manipolare (analisi e interpretazione): cosa è possibile fare con questi dati? I dati devono essere analizzati, interrogati, devono essere capaci di restituire conoscenza una volta che siano diventati un'informazione manipolabile da parte della macchina. Questo significa tanto ragionare a livello di costruzione di un'infrastruttura per ospitare le applicazioni quanto nella creazione di applicazioni ad hoc capaci di restituire il massimo potere conoscitivo espresso in potenza dalle risorse, espresse attraverso i dati.
- Archiviare (conservazione): la preservazione. L'archiviazione è un tema che riguarda sicuramente il servizio di *storage*, ma che richiama anche l'importante tema dei metadati di preservazione. E non va dimenticato che è certamente necessario mettere in atto politiche di preservazione dei dati, ma in molti casi è necessario garantire anche la preservazione dell'applicazione (e talune volte dell'intero ambiente) che è in

grado di leggere quei dati per restituirli ad un fruitore. Sempre di più la preservazione dei dati avviene su *cloud*, ma è necessario ricordare che bisogna disporre di infrastrutture in grado di ospitare dati e applicazioni per la loro accessibilità nel tempo (certi dati sono intellegibili solo quando ci sia l'applicazione adatta a interpretarli).

- Divulgare (disseminazione): accesso e *design* di interfaccia. È innegabile che in questi ultimi anni ci sia stato un cambio di cultura visiva delle interfacce e di questo va tenuto conto in fase di *design* lato utente. E poi c'è un altro importante tema, che è quello della differenziazione delle viste, della presa di coscienza del fatto che non tutti gli utenti sono interessati agli stessi contenuti. La tradizione dell'accesso profilato ai contenuti Web si traduce in una differenziazione della vista sui dati che è proporzionale ai quesiti di ricerca posti.
- Usare e riusare. Facilitare l'accesso e garantire il riutilizzo di dati e applicazioni sviluppati in seno ad un progetto, ovvero anche garantire la vita ai propri dati in contesti diversi da quello nativo.

Questa serie di azioni ci fanno capire che ogni step richiede soluzioni (e strumenti) ad hoc seppure in un circuito integrato. E a questo proposito particolarmente rilevante è il tema della *digital curation*, che potremmo tradurre in un interessante concetto relativo al ruolo di chi esercita questo mestiere: il curatore "è un 'produttore di significato' o almeno una persona che facilita la produzione di significato degli oggetti digitali su cui opera".[296]

Se dovessimo cercare di definire il concetto di *digital curation* ci dovremmo confrontare con la definizione proposta dal Digital Curation Centre (DCC): "Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle".[297] Va osservato come questa definizione assegna la *digital curation* ad una prospettiva dinamica, che fa riferimento non tanto e non solo ai singoli passi che riguardano la gestione degli oggetti digitali, quanto piuttosto all'intero ciclo di vita di essi.[298]

Riprendendo ancora le parole di Laura Testoni, il ciclo di vita della *digital curation*, secondo il DCC, può essere riassunto in queste fasi:

- concettualizzazione: la creazione di dati digitali va prima pianificata, includendo nella pianificazione i metodi di cattura e le opzioni di immagazzinamento (*storage*);
- creazione o ricezione: è la fase in cui il dato viene prodotto, o ricevuto "grezzo" dall'autore e comporta l'assegnazione dei diversi metadati (si noti: senza l'assegnazione di metadati non c'è creazione di dati);
- valutazione e selezione: è necessario selezionare gli oggetti digitali che necessiteranno di una conservazione a lungo termine, in ottemperanza a linee guida, regole e requisiti legali;
- acquisizione (*ingest*): è il momento in cui si trasferiscono gli oggetti digitali in un archivio o in un *repository*;
- azione di conservazione: si tratta di assicurare la conservazione a lungo termine e la preservazione dell'integrità (*retention of the authoritative nature*) dei dati;
- *store*: i dati devono essere conservati in modo sicuro e aderente agli standard;
- accesso e uso e riutilizzo: è necessario assicurarsi che i dati siano accessibili per l'uso quotidiano da parte degli utenti designati;
- trasformazione: la creazione di nuovi dati dagli originali, sia attraverso una rielaborazione (ad esempio creazione di subset) sia, eventualmente, attraverso migrazione in un differente formato.[299]

Ecco che la *digital curation* è un concetto che va oltre la pura preservazione del dato, per abbracciare ogni momento del processo che fa del dato uno strumento di conoscenza. E la *curation* andrà ripensata in un'ottica che certo valorizzi il dato, ma che sia consapevole della presenza di un circuito: *curation* dei dati, *curation* dell'applicazione che legge quei dati e *curation* del supporto che conserva quei dati.

Per concludere. Date queste riflessioni capiamo che ogni step del nostro ciclo vitale ha problemi di: scelta dei formati più adeguati alla rappresentazione, definizione degli strumenti necessari a gestire gli oggetti prodotti, selezione dei modelli più appropriati a garantire l'espressività del contenuto di quegli oggetti, acquisizione dei metodi necessari a manipolare formati, dati e oggetti e infine creazione o riuso di tools adeguati rispetto ai bisogni. Diremo che i problemi del ciclo vitale corrispondono a vari profili, diremo, professionali o anche a nuove identità capaci di sovrintendere ad un processo articolato e multiforme. Se dovessimo immaginare il mestiere del *digital humanist* dovremmo allora pensare nei termini di

un percorso che dall'elaborazione di un'idea debba portare alla narrazione di una storia. E chiudiamo allora, senza alcuna pretesa di esaustività, con una prima proposta di profili di esperto in DH, che, con gli strumenti del Web semantico, sappia far fronte a problemi di organizzazione della conoscenza, in particolare nel contesto di archivi, biblioteche e musei. Profili che qualificano le nuove figure del futuro,[300] e che naturalmente emergono dall'esperienza condotta dentro a DHDK, e in un più ampio contesto di offerta formativa nel settore delle DH in Italia:[301]

- *Metadata Specialist*. Il professionista che stabilisce il set di metadati più adeguato in base alle caratteristiche della raccolta, agli obiettivi del progetto e agli scopi della ricerca (il primato della descrizione nelle azioni di processo);
- *Content strategist*. Il professionista che lavora sul testo e che valorizza l'importanza del contenuto (il markup, l'edizione di un testo o l'edizione delle fonti primarie) ed è quindi anche consapevole del dominio di riferimento. In altri contesti, può assumere l'identità del *Web Content Specialist*, cioè di chi governa i problemi di rappresentazione del contenuto di un oggetto anche in ambiente distribuito;
- *User interface designer* o *User Experience Designer*. Il professionista che ragiona sull'importanza dell'utente. Questo significa anche conoscere il Web e i problemi di usabilità delle interfacce; può rientrare in questa definizione anche il *Web Information Architect*, come colui o colei capace di dialogare con chi deve rispondere all'esigenza di tradurre l'analisi dei dati in sistemi anche visuali (*Web analyst*);
- *Knowledge Engineer*. Il professionista che lavora sulla componente semantica (Semantic Web e LOD), esperto di come costruire ontologie o in generale strumenti concettuali di rappresentazione della conoscenza e si affianca al *Knowledge organizer* in senso stretto come colui o colei che gestisce tale semantica anche per fornire un servizio all'utente;
- *Multimedia manager* o *Multimedia Object specialist*. Il professionista che va oltre il testo (mappe interattive, oggetti non testuali, 3D, nuovi media) e che può anche avere le competenze necessarie alla progettazione di nuovi ambienti di fruizione di tipologie diverse di formati;

- *Data curator*. Il professionista che si prende cura dell'intero ciclo vitale delle risorse digitali, che conosce il *workflow* necessario alla valorizzazione del dominio;
- *Digital Publishing Expert*. Il professionista che governa il tema dell'editoria digitale e che conosce i principi a fondamento dei dati aperti, anche sul piano giuridico, in un più generale contesto di open science;
- *Digital Library Specialist*. Unisce le competenze degli altri profili ad un livello di *project managing*: architettura di sistema, modelli di produzione di dati, creazione di pacchetti di gestione e conservazione, sistemi di disseminazione, strumenti e servizi di accesso.

E siamo così arrivati davvero alle conclusioni.

Dalla computabilità delle discipline umanistiche all'esigenza di valorizzare l'atto interpretativo dello studioso come espressione di un punto di vista. Dall'importanza dei contesti come strumento a garanzia dell'espressività dei dati, al ragionamento a livelli attraverso cui osservare gli oggetti digitali. Dal lavoro a progetto al ciclo vitale delle risorse digitali. Ecco che si delinea una nuova identità di studioso che, con gli strumenti del Web semantico, contribuisce a consegnare la memoria analogica a nuove esperienze di accesso e fruizione, contribuisce a disegnare nuovi servizi che sfruttano la capacità dei dati di dialogare in un sistema a rete, trasversale e pluridirezionale. Ecco che le DH, assieme al Web semantico, propongono nuovi modelli di organizzazione della conoscenza in un sistema archivi, biblioteche e musei non più rappresentati come *silos* di dati, ma come entità in comunicazione, organismi dotati di quel comune obiettivo che è la realizzazione di nuove istanze esperienziali per un pubblico vasto ed eterogeno. Ecco le nuove frontiere del sapere. Ecco il ruolo delle DH nel disegnare il *knowledge graph* del patrimonio culturale come insieme di dataset semanticamente espressivi e profondamente relazionati, capaci di assolvere all'intera filiera della gestione della conoscenza (Fig. 14).

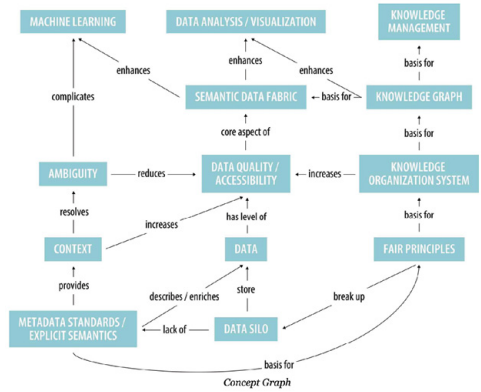

Figura 14 - Un *Concept Graph* per presentare gli elementi chiave del progetto di costruzione di reti di conoscenza.
Tratto da Andreas Blumauer, Helmut Nagy, *The Knowledge Graph Cookbook*, Kindle format, 2020

284 Alcuni repertori di progetti: per l'Italia la sezione progetti del sito AIUCD, http://www.aiucd.it/progetti, che ha il suo corrispettivo in EADH, https://eadh.org/projects.
285 Si veda in particolare la raccolta dei progetti realizzati dagli studenti nel dataset DHDKey!, https://projects.dharc.unibo.it/dhdkey.
286 Per una descrizione puntuale della struttura del corso di studi si veda Francesca Tomasi, *DIGITAL HUMANITIES AND DIGITAL KNOWLEDGE (DHDK). International second cycle/Master degree*, "Umanistica Digitale", 2 (2018).
287 Johanna Drucker, *DH101. INTRODUCTION TO DIGITAL HUMANITIES. Concepts, Methods, and Tutorials for Students and Instructors*, 2014, http://dh101.humanities.ucla.edu. Purtroppo a oggi la risorsa non è più accessibile da questo URL, ma si spera che a breve il sito torni a essere disponibile.
288 Fra i progetti finalizzati a classificare azioni, prodotti, principi e metodi nelle DH si possono vedere: TaDiRAH, http://tadirah.dariah.eu/vocab; CLARIN Concept Registry Browser, https://concepts.clarin.eu/ccr/browser; Tapor3, http://tapor-test.artsrn.ualberta.ca/home; The NeDiMAH-DARIAH research methods ontology project (NeMO) – anche se il sito ufficiale pare a oggi non attivo, per informazioni si può leggere https://dh2016.adho.org/abstracts/134.
289 Interessante il blog di Miriam Posner che dedica al tema *How did they make that?* una sezione, in cui decostruisce alcuni progetti di DH ("a gallery

of primary sources, a digital scholarly edition, a mapping project, a network visualization, computer-aided text analysis, a historical 3D model, a longform media-rich narrative"), andando ad analizzare linguaggi, tool, strumenti, librerie, software utilizzati per realizzare i progetti, https://miriamposner.com/blog/how-did-they-make-that.

290 "RIDE is a peer-reviewed academic journal published by the Institut für Dokumentologie und Editorik dedicated to digital editions and resources". Pagina ufficiale di RIDE, https://www.i-d-e.de/publikationen/ride.

291 Si tratta di tre linee guida leggermente differenti sulla base della tipologia di risorsa da recensire (edizioni, collezioni di testi, tool e ambienti): *Criteria for Reviewing Scholarly Digital Editions*; *Criteria for Reviewing Text Collections*; *Criteria for Reviewing Tools and Environments for Digital Scholarly Editing*. Si veda la pagina aggregata con i link alle linee guida, riportate in diverse lingue, https://ride.i-d-e.de/reviewers/catalogue-criteria-for-reviewing-digital-editions-and-resources.

292 Per un primo approccio al tema FAIR, si veda https://www.go-fair.org/fair-principles; in particolare si può leggere Mark D. Wilkinson, et al., *The FAIR Guiding Principles for scientific data management and stewardship*, "Scientific Data", 3 (2016), https://www.nature.com/articles/sdata201618.

293 María Poveda-Villalón, Daniel Garijo, *Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web*, 2020, https://arxiv.org/pdf/2003.13084.pdf; María Poveda-Villalón, Paola Espinoza-Arias, Daniel Garijo, *Coming to Terms with FAIR Ontologies*, in *Knowledge Engineering and Knowledge Management*, LNCS 12387, Berlin, Springer, 2020. Si veda anche Francesco Beretta, *A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)*, "Semantic Web", 12 (2021), 2, p. 279-294, special issue on Semantic Web for Cultural Heritage, https://doi.org/10.3233/SW-200416.

294 Data Catalog Vocabulary (DCAT), https://www.w3.org/TR/vocab-dcat.

295 Sebastian Barzaghi, *Valutazione comparativa per la progettazione e lo sviluppo dell'Edizione Nazionale delle Opere di Aldo Moro*, Report tecnico, Zenodo, 2021, https://doi.org/10.5281/zenodo.5213156.

296 Laura Testoni, *Digital curation e content curation: due risposte alla complessità dell'infosfera digitale che ci circonda, due sfide per i bibliotecari*, "Bibliotime", 16 (2013), 1.

297 Si veda il *Curation Lifecycle Model* del Digital Curation Centre (DCC), https://www.dcc.ac.uk/guidance/curation-lifecycle-model

298 Un modello grafico del *DCC Curation Lifecycle Model* si può consultare alla pagina https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf.

299 Laura Testoni, *Digital curation e content curation*, cit.

300 Interessante consultare lo statunitense Code4Lib alla sezione relativa ai post sui lavori legati al dominio MAB/(G)LAM, https://jobs.code4lib.org. Ma si veda anche la sezione Job announcement su DHnow, https://

digitalhumanitiesnow.org.
301 Vale la pena citare almeno il rapporto Almalaurea 2021 sulla condizione occupazionale dei laureati in DH, https://www.almalaurea.it/informa/news/2021/07/22/digital-humanities-e-contaminazioni-disciplinari.

BIBLIOGRAFIA

A Companion to Digital Humanities, edited by Susan Schreibman, Ray Siemens, John Unsworth, Oxford, Blackwell, 2004, http://www.digitalhumanities.org/companion.

Ackoff Russell, *From Data to Wisdom*, "Journal of Applied Systems Analysis", 16 (1989), p. 3-9.

Adamo Giovanni, *Bibliografia di Informatica umanistica*, Roma, Bulzoni, 1994.

Agenzia per l'Italia Digitale (AGID), Commissione di Coordinamento SPC, *Linee guida per l'interoperabilità semantica attraverso i Linked Open Data*, versione 2.0, 2012, https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf.

Id., *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, ultima release 2020, https://docs.italia.it/italia/daf/lg-patrimonio-pubblico/it/stabile/index.html.

AIUCD 2019, *Book of Abstracts*, Quaderni di Umanistica Digitale, 2019, https://doi.org/10.6092/unibo/amsacta/6361.

Almanacco letterario Bompiani 1962: le applicazioni dei calcolatori elettronici alle scienze morali e alla letteratura, a cura di Sergio Morando, Milano, Bompiani, 1961.

Anderson Sheila, Blanke Tobias, Dunn Stuart, *Methodological commons: arts and humanities e-Science fundamentals*, "Philosophical Transactions of The Royal Society. A Mathematical Physical and Engineering Sciences", 368 (2010), p. 3779–3796, https://doi.org/10.1098/rsta.2010.0156.

Ansani Michele, *Edizione digitale di fonti diplomatiche: esperienze, modelli testuali*, "RM Reti Medievali", 7 (2006), 2, https://doi.org/10.6092/1593-2214/140.

Associazione Italiana Biblioteche (AIB), *La cooperazione: il Servizio bibliotecario nazionale*, a cura di Antonella Aquilina D'Amore. Atti del 30 congresso nazionale dell'Associazione italiana biblioteche (Giardini-Naxos, 21-24 novembre 1982), Messina, Università di

Messina, Facoltà di lettere e filosofia, Centro studi umanistici, 1986.
Authority control: definizione ed esperienze internazionali: atti del convegno internazionale (Firenze, 10-12 febbraio 2003), a cura di Mauro Guerrini, Barbara B. Tillett, Firenze, Firenze University Press; Roma, AIB, 2003.
Banerjee Kyle, *The Linked Data Myth*, "Library Journal", News (August 13, 2020), https://bit.ly/2VIXPnQ.
Barabucci Gioele, Tomasi Francesca, Vitali Fabio, *Supporting Complexity and Conjectures in Cultural Heritage Descriptions*, in *COLCO 2020. Collect and Connect: Archives and Collections in a Digital Age* 2020, "CEUR Workshop Proceedings", 2810 (2021), p. 104-115, http://ceur-ws.org/Vol-2810.
Id., *Modeling data complexity in public history and cultural heritage*, in *Handbook Digital Public History*, edited by Serge Noiret, Mark Tebeau, Gerben Zaagsma, Berlin, De Gruyter, 2022 (in corso di stampa).
Barzaghi Sebastian, *La modellazione dei dati nell'Edizione Nazionale delle Opere di Aldo Moro,* Zenodo, 2021, https://doi.org/10.5281/zenodo.5144961.
Id., *Valutazione comparativa per la progettazione e lo sviluppo dell'Edizione Nazionale delle Opere di Aldo Moro*, Report tecnico, Zenodo, 2021, https://doi.org/10.5281/zenodo.5213156.
Beretta Francesco, *A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)*, "Semantic Web", 12 (2021), 2, p. 279-294, special issue on Semantic Web for Cultural Heritage, https://doi.org/10.3233/SW-200416.
Berners-Lee Tim, Fischetti Mark, *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor*, San Francisco, Harper San Francisco, 1999.
Berners-Lee Tim, Hendler James, Lassila Ora, *The Semantic Web*, "Scientific American", 284 (2001), 5, p. 34-43.
Berners-Lee Tim, *Linked Data*, 2006, https://www.w3.org/DesignIssues/LinkedData.html.
Bianchini Carlo, *Organizzare la conoscenza con la sequenza di filiazione della Classificazione Colon di S.R. Ranganathan*, "JLIS.it", 2 (2011), 2, https://doi.org/10.4403/jlis.it-4710.
Bianchini Carlo, Guerrini Mauro, *Introduzione a RDA. Linee guida per rappresentare e scoprire le risorse*, Milano, Editrice Bibliografica,

2014.
Bianconi Lorenzo, Pompilio Angelo, Pagannone Gennaro, *RADAMES: prototipo d'un repertorio e archivio digitale per il melodramma*', "Il Saggiatore Musicale", 11 (2004), 2, p. 345-394.
Bliss Henry E., *A modern classification for libraries, with simple notation, mnemonics, and alternatives*, "Library Journal", 35 (1910), p. 351-358.
Blumauer Andreas, Nagy Helmut, *The Knowledge Graph Cookbook*, Kindle format, 2020.
Bonora Paolo, Pompilio Angelo, *"Osservate, leggete con me". Risorse LOD per la storia del melodramma: una prospettiva funzionale di rappresentazione*, in *AIUCD 2019*, *Book of Abstracts*, Quaderni di Umanistica Digitale, 2019, p. 202-2017, https://doi.org/10.6092/unibo/amsacta/6361.
Bonora Paolo, Pompilio Angelo, *Corago in LOD. The debut of an Opera repository into the Linked Data arena*, "JLIS.It", 12 (2021), 2, p. 54–72, https://doi.org/10.4403/jlis.it-12699.
Borko Harold, *Information science: What is it?*, "American Documentation", 19 (1968), 1, p. 3-5.
Bradley John, Pasin Michele, *Factoid-based prosopography and computer ontologies: towards an integrated approach*, "Digital Scholarship in the Humanities", 30 (2015), 1, p. 86-97, http://doi.org/10.1093/llc/fqt037.
Bruni Silvia et al., *Verso l'integrazione tra archivi, biblioteche e musei. Alcune riflessioni*, "JLIS.it", 7 (2016), 1, p. 225-44, https://doi.org/10.4403/jlis.it-11482.
Busa Roberto, *The annals of humanities computing: The index thomisticus,* "Computers and the Humanities", 14 (1980), 2, p. 83-90.
Bush Vannevar, *As We May Think*, "The Atlantic Monthly", 176 (1945), 1, p. 101-108.
Buttò Simonetta, *Alphabetica, il nuovo portale per la ricerca integrata: un salto di qualità per le biblioteche italiane*, "DigItalia. Rivista del digitale nei beni culturali", 14 (2019), 1, p. 9-28, https://doi.org/10.36181/digitalia-00010.
Buzzetti Dino, *Digital representation and the text model*, "New Literary History", 33 (2002), p. 61-88.
Buzzetti Dino, McGann Jerome, *Critical editing in a digital horizon*, in *Electronic textual editing,* edited by John Unsworth, Katherine O'Brien O'Keeffe, Lou Burnard, New York, MLA of America,

2006, p. 53-73.
Candela Gustavo et al., *Evaluating the Quality of Linked Open Data in Digital Libraries*, "Journal of Information Science", (August 2020), https://doi.org/10.1177/0165551520930951.
Carriero Valentina Anita, Daquino Marilena, Tomasi Francesca, *Convergenze semantiche tra musei, archivi e biblioteche. Ontologie per le relazioni interpersonali,* "JLIS.it", 10 (2019), 1, p. 72-91.
Carroll Jeremy J., Bizer Christian, Hayes Pat, Stickler Patrick, *Named graphs*, "Journal of Web Semantics", 3 (2005), 4, p. 247-267.
Id., *Named graphs, provenance and trust*, in *Proceedings of the 14th international conference on World Wide Web (WWW '05)*, NY, ACM, 2005, p. 613-622, https://doi.org/10.1145/1060745.1060835.
Codd Edgar F., *A relational model of data for large shared data banks*, "Communications of the ACM", 13 (1970), 6, p. 377-387.
Computing in Humanities Education: A European Perspective, ACO*HUM, edited by Koenraad De Smedt et al., University of Bergen, HIT-centre, 1999, http://korpus.uib.no/humfak/AcoHum/book.
Consultative Committee for Space Data Systems, *Reference Model For An Open Archival Information System (OAIS)*, 2012, https://public.ccsds.org/pubs/650x0m2.pdf.
Coyle Karen, *FRBR Before and After*, Chicago, ALA, 2016; trad. it. *FRBR prima e dopo*, a cura di Lucia Sardo, AIB, 2017.
Daquino Marilena, *Mining Authoritativeness in Art Historical Photo Archives: Semantic Web Applications for Connoisseurship*, Studies on the Semantic Web 40, Berlin, IOS Press, 2019.
Daquino Marilena, Peroni Silvio, Tomasi Francesca, Vitali Fabio, *Political Roles Ontology (PRoles): enhancing archival authority records through Semantic Web technologies*, "Procedia Computer Science", 38 (2014), Elsevier, p. 60-67.
Daquino Marilena, Tomasi Francesca, *Digital Humanities e Library and Information Science. Through the lens of knowledge organization*, "Bibliothecae.it", 5 (2016), 1, p. 130-150, https://doi.org/10.6092/issn.2283-9364/6109.
Daquino Marilena et al., *Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data*, "ACM Journal on Computing and Cultural Heritage" (JOCCH), 10 (2017), 4, p. 1-21, http://

dx.doi.org/10.1145/3051487.
Daquino Marilena, Giovannetti Francesca, Tomasi Francesca, *Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini*, "Umanistica Digitale", 3 (2019), 7, https://doi.org/10.6092/issn.2532-8816/9091.
Daquino Marilena, Pasqual Valentina, Tomasi Francesca, *Knowledge Representation of digital Hermeneutics of archival and literary Sources*, "JLIS.it", 11 (2020), 3, p. 59-76, http://dx.doi.org/10.4403/jlis.it-12642.
De Gennaro Richard, *Library automation & networking perspectives on three decades*, "Library journal", 108 (1983), 7, p. 629-635.
Defining digital humanities: a reader, edited by Melissa Terras, Julianne Nyhan, Edward Vanhoutt, London, Routledge, 2013.
DeRose Steven J., Durand David G., Mylonas Elli, Renear Allen H., *What is Text, Really?*, "Journal of Computing in Higher Education", 1 (1990), 2, p. 3-26.
Dillon Andrew, *What It Means to Be an ISchool*, "Journal of Education for Library and Information Science", 53 (2012), 4, p. 267-273.
Doerr Martin, *Ontologies for Cultural Heritage*, in *Handbook on Ontologies*, edited by Steffen Staab, Rudi Studer, Berlin, Springer, 2009, p. 463-486.
Drucker Johanna, *Humanities Approaches to Graphical Display*, "Digital Humanities Quarterly", 5 (2011), 1, http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.
Ead., *DH101. INTRODUCTION TO DIGITAL HUMANITIES. Concepts, Methods, and Tutorials for Students and Instructors*, 2014, http://dh101.humanities.ucla.edu.
Ehrlinger Lisa, Wöß Wolfram, *Towards a definition of knowledge graphs*, "SEMANTICS" (Posters & Demos Track), 48 (2016), p. 1-4.
Electronic Textual Editing, edited by John Unsworth, Katherine O'Brien O'Keeffe, Lou Burnard, New York, MLA of America, 2006.
Elings Mary W., Waibel Gunter, *Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives and Museums*, "First Monday", 12 (2007), 3, http://firstmonday.org/issues/issue12_3/elings/index.html.
Encyclopedia of Library and Information Sciences, edited by John D.

McDonald, Michael Levine-Clark, Fourth edition Vol. 1-7, Boca Raton, FL, CRC Press, 2017.

Factor Michael, et al., *Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware Storage,* "First workshop on the Theory and Practice of Provenance", 2009, 6, p. 1-10, https://dl.acm.org/doi/10.5555/1525932.1525938.

Fagnoni Enrico et al., *How to use Linked Data*, 2015, https://en.wikitolearn.org/Course:How_to_use_Linked_Data/Interaction_with_Linked_Data/Linked_Data_Visualization

Feliciati Pierluigi, *Archives in a Graph. The Records in Contexts Ontology within the framework of standards and practices of Archival Description*, "JLIS", 12 (2021), 1, p. 92-101, http://dx.doi.org/10.4403/jlis.it-12675

Fiormonte Domenico, *Scrittura e filologia nell'era digitale*, Torino, Bollati Boringhieri, 2003.

Flanders Julia, Jannidis Fotis, *Knowledge Organization and Data Modeling in the Humanities*, [White Paper], 2015, https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/11127.

Froger Jacques, *La critique des textes et son automatisation*, Paris, Dunod, 1968.

Id., *La méthode de Dom Quentin, la méthode des distances et le problème de la contamination*, in *La pratique des ordinateurs dans la critique des textes*, a cura di Jean Irigoin, Gian Piero Zarri, Colloque International du CNRS nr. 579, Paris, Editions du CNRS, 1979, p. 13-22.

Gavin Michael, *Vector Semantics, William Empson, and the Study of Ambiguity*, "Critical Inquiry", 44 (2018), 4, p. 641-673, https://web.stanford.edu/~jurafsky/slp3/6.pdf.

Giovannetti Francesca, Tomasi Francesca, *The Linked Finding Aid as a Platform for Textual Research: The Case Study of the Giuseppe Raimondi Archive*, in *Linked Archives 2021*, in *Proceedings of Linked Archives*, International Workshop 2021, co-located with 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021), "CEUR Workshop Proceedings", 3019 (2021), p. 104-113, http://ceur-ws.org/Vol-3019.

Giuseppe Raimondi: carte, libri, dialoghi intellettuali, Bologna, Patron, 1998.

Gnoli Claudio, Marino Vittorio, Rosati Luca, *Organizzare la*

conoscenza: dalle biblioteche all'architettura dell'informazione per il Web, Pavia, Hops, 2006.

Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing, edited by Elena Spadini, Francesca Tomasi, Georg Vogeler, Norderstedt, Books on Demand, 2021.

Groth Paul, Gibson Andrew, Velterop Jan, *The Anatomy of a Nanopublication*. "Information Services & Use", 30 (2010), 1-2, p. 51-56, https://doi.org/10.3233/ISU-2010-0613.

Hartig Olaf, *Foundations of RDF* and SPARQL*: an alternative approach to statement-level metadata in RDF*, in *11th Alberto Mendelzon International Workshop on Foundations of Data Management (AMW),* "CEUR Workshop Proceedings", 1912 (2017), http://ceur-ws.org/Vol-1912.

Hjørland Birger, *Library and information science: practice, theory, and philosophical basis* "Information Processing and Management", 36 (2000), p. 501-531.

Id., *What is Knowledge Organization (KO)?*, "Knowledge Organization", 35 (2008), 2, p. 86-101, https://doi.org/10.5771/0943-7444-2008-2-3-86.

Id., *Information Science and Its Core Concepts: Levels of Disagreement*, in *Theories of Information, Communication and Knowledge*, edited by Fidelia Ibekwe-SanJuan, Thomas Dousa, Studies in History and Philosophy of Science 34, Dordrecht, Springer, 2014, p. 205-235, https://doi.org/10.1007/978-94-007-6973-1_9.

Hockey Susan, *Electronic texts in the humanities: principles and practice*, Oxford, Oxford University Press, 2000.

Hodge Gail, *Systems of knowledge organization for digital libraries. Beyond traditional authority files*, Washington, DC, The Council on Library and Information Resources, 2020, http://www.clir.org/pubs/reports/pub91/contents.html.

Hyvönen Eero, *Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery*, "Semantic Web", 11 (2020), 1, p. 187-193, https://doi.org/10.3233/SW-190386.

Un'indagine sui programmi di inventariazione archivistica, "Archivi & Computer", 13 (2003), 3.

ICA/CDS, *Linee-guida per l'elaborazione e la presentazione di strumenti di ricerca*, trad. it. 2004, a cura di Francesca Ricci, http://media.regesta.com/dm_0/ANAI/anaiCMS/ANAI/000/0111/

ANAI.000.0111.0007.pdf.
Isaac Antoine, Manguinhas Hugo, Stiller Juliane, Charles Valentine, *Report on enrichment and evaluation*, The Hague, Netherlands, Europeana Task Force on Enrichment and Evaluation, 2015, https://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf.
Italia Paola, *Editing Duemila. Per una filologia dei testi digitali*, Roma, Salerno Editrice, 2020.
Jakobson Roman, *Linguistics and Poetics*, in *Style in language*, edited by Thomas A. Sebeok, New York, John Wiley & Sons, 1960, p. 350-377.
Lana Maurizio, *Biblioteche digitali. Un'introduzione*, Bologna, Bononia University Press, 2013.
Id., *Digital humanities e biblioteche*, "AIB studi", 59 (2019), 1-2, https://doi.org/10.2426/aibstudi-11862.
Landow George P., *L'ipertesto: tecnologie digitali e critica letteraria*, Milano, Bruno Mondadori, 1998.
Linee guida sul trattamento dei fondi personali, a cura della Commissione nazionale biblioteche speciali, archivi e biblioteche d'autore, versione 15.1, 2019, https://www.aib.it/wp-content/uploads/2020/01/15.1_Linee-Guida-fondi-personali-def-1-ULTIMA-VERSIONE.pdf.
Llanes-Padrón Dunia, Pastor-Sánchez Juan-Antonio, *Records in contexts: the road of archives to semantic interoperability*, "Program", 51 (2017), 4, p. 387-405, https://doi.org/10.1108/PROG-03-2017-0021.
Mancinelli Maria Letizia, *Gli standard catalografici dell'Istituto Centrale per il Catalogo e la Documentazione,* in Roberta Tucci, *Le voci, le opere e le cose. La catalogazione dei beni culturali demoetnoantropologici*, Roma, ICCU, 2018, p. 279-302.
Mazzini Silvia, Ricci Francesca, *EAC-CPF Ontology and Linked Archival Data*, in *Proceedings of the 1st International Workshop on Semantic Digital Archives (SDA 2011)*, "Ceur Workshop Proceedings", 801 (2011), p. 72-81, http://ceur-ws.org/Vol-801/paper6.pdf.
McCarty Willard, Short Harold, *Mapping the field*, Report of ALLC meeting held in Pisa, April 2002, http://www.allc.org/node/189.
McCarty Willard, *Humanities Computing*, London and New York,

Palgrave, 2005.
Meschini Federico, *Oltre il libro. Forme di testualità e digital humanities*, Milano, Editrice Bibliografica, 2020.
Metitieri Fabio, Ridi Riccardo, *Biblioteche in rete. Istruzioni per l'uso*, Roma-Bari, Laterza, 2006.
Michetti Giovanni, *Ma è poi tanto pacifico che l'archivio rispecchi l'albero?*, "Archivi e Computer", 14 (2009), 1, p. 85-95.
Id., *Archives are not a Tree*, in *The Memory of the World in the Digital Age: Digitization and Preservation*, edited by Luciana Duranti, Elizabeth Shaffer, Vancouver, UNESCO, 2013, p. 1002-1010.
Moretti Franco, *Conjectures on World Literature*, "New Left Review", 1 (2000).
Morgan Eric Lease and LiAM, *Linked Archival Metadata: A Guidebook*, version 0.99, 2014, http://infomotions.com/sandbox/liam/tmp/guidebook.pdf.
Moro Aldo, *Edizione Nazionale delle Opere di Aldo Moro*, Bologna, Università di Bologna, 2021, https://doi.org/10.6092/unibo/aldomoro.
Neill Lian, *Building a Graph of History with The Codex*, 2020, https://neo4j.com/blog/building-graph-history-codex.
Neill Lian, Schmidt Desmond, *SPEEDy. A Practical Editor for Texts Annotated with Standoff Properties*, in *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, Georg Vogeler, Norderstedt, Books on Demand, 2021, p. 45-54.
Nelson Theodor H., *A File Structure for The Complex, The Changing and the Indeterminate*, in *Proceedings of ACM 20th National Conference*, New York, ACM, 1965, p. 84-100.
Niu Jinfang, *Linked Data for Archives*, "Archivaria", 82 (2016), p. 83-110, https://archivaria.ca/index.php/archivaria/article/view/13582.
Numerico Teresa, *Alan Turing e l'intelligenza delle macchine*, Milano, Franco Angeli, 2004.
Nyhan Julianne, Flinn Andrew, *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Springer Open, 2018.
O' Reilly Tim, *What is Web 2.0.*, 2005, http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.
Orlandi Tito, *Informatica Umanistica*, Roma, La Nuova Italia Scientifica, Studi superiori 78, 1990.

Otlet Paul, *Traité de documentation. Le livre sur le livre. Théorie et pratique*, Bruxelles, Editiones Mundaneum, 1934.

Paolo Bufalini, Appunti (1981-1991), a cura di Marilena Daquino, Martina Dello Buono, Francesca Giovannetti, Francesca Tomasi, Bologna, Digital Humanities Advanced Research Centre (/DH.arc), Università di Bologna, 2020, https://doi.org/10.6092/unibo/amsacta/6415.

Pasqual Valentina, Daquino Marilena, Tomasi Francesca, *MIMA: A Data Model to Represent Multi-Disciplinary Analysis on Manuscripts. Use Case on Pellegrino Prisciani's Historiae Ferrariae*, in *AIUCD 2021, Book of extended abstracts*, Quaderni di Umanistica Digitale, p. 560-566, https://doi.org/10.6092/unibo/amsacta/6712.

Passarotti Marco, Rockwell Geoffrey, *The Index Thomisticus as a Big Data Project,* "Umanistica Digitale", 5 (2019), http://doi.org/10.6092/issn.2532-8816/8575.

Pattuelli Cristina, *Mapping People-centered Properties for Linked Open Data*, "Knowledge Organization", 38 (2011), 4, p. 352-359.

Pellizzi Federico, *Letteratura biblioteche ipertesti*, Roma, Carocci, 2005.

Perilli Lorenzo, *Filologia Computazionale*, Roma, Accademia Nazionale dei Lincei, 1995.

Peruginelli Susanna, et al., *RDA e archivi: ricerca di un raccordo tra mondi diversi*, "JLIS.it", 9 (2018), 1, p. 137-147, https://dx.doi.org/10.4403/jlis.it-12403.

Pierazzo Elena, *A rationale of digital documentary editions*, "Literary and Linguistic Computing", 26 (2011), 4, p. 463-477, https://doi.org/10.1093/llc/fqr033.

Ead., *Digital Scholarly Editing: Theories, Models and Methods*, Londra, Routledge, 2015.

Piotrowski Michael, *Accepting and Modeling Uncertainty*, "Zeitschrift für Digitale Geisteswissenschaften", 2019, http://doi.org/10.17175/sb004_006a.

Pitti Daniel, *Creator Description: Encoded Archival Context*, in *Authority control in organizing and accessing information: definition and international experience*, edited by Arlene G. Taylor, et al., Binghamton N.Y., Haworth Information Press, 2004, p. 201-226.

Po Laura, Bikakis Nikos, Desimoni Federico, Papastefanatos George, *Linked Data Visualization: Techniques, Tools, and Big Data*, "Synthesis Lectures on the Semantic Web: Theory and

Technology", 10 (2020), 1, https://doi.org/10.2200/S00967ED1V01Y201911WBE019.
Poveda-Villalón María, Espinoza-Arias Paola, Garijo Daniel, *Coming to Terms with FAIR Ontologies*, in *Knowledge Engineering and Knowledge Management*, LNCS 12387, Berlin, Springer, 2020.
Poveda-Villalón María, Garijo Daniel, *Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web*, 2020, https://arxiv.org/pdf/2003.13084.pdf.
La pratique des ordinateurs dans la critique des textes, a cura di Jean Irigoin, Gian Piero Zarri, Colloque International du CNRS nr. 579, Paris, Editions du CNRS, 1979.
Quentin Henri, *Essais de critique textuelle (Ecdotique)*, Parigi, Picard, 1926.
Ranganathan Shiyali Ramamrita, *Colon Classification* [First edition], Madras, Madras Library Association, 1933.
Records in Contexts. A conceptual model for archival description. Il contributo italiano, "Il mondo degli archivi", Quaderni (2017), 2, http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n2.pdf.
Renear Allen H., Mylonas Elli, Durand David G., *Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies*, prima versione 1993, seconda versione comparsa nella serie *Research in Humanities Computing* 4, edited by Nancy Ide and Susan Hockey, Oxford, Oxford University Press, 1996, p. 263-280.
Riley Jenn, *Understanding metadata. What is metadata and what is it for?*, a Primer Publication of the National Information Standards Organization (NISO), USA, 2017, https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf.
Riva Pat, *Il nuovo modello concettuale dell'Universo bibliografico: FRBR Library Reference Model*, "AIB studi", 56 (2016), 2, p. 265-275.
Robinson Helena, *Remembering things differently: Museums, libraries and archives as memory institutions and the implications for convergence*, "Museum Management and Curatorship", 27 (2012), 4, p. 413-429.
Salarelli Alberto, *International Image Interoperability Framework (IIIF): a panoramic view*, "JLIS", 8 (2017), 1, p. 50-66, https://dx.doi.org/10.4403/jlis.it-12090.
Santoro Michele, *Ripensare la CDU. Per una riflessione sulla storia, il*

ruolo e le prospettive della Classificazione Decimale Universale, "Biblioteche oggi", 13 (1995), 8, p. 48-57.

Seeing Standards: A Visualization of the Metadata Universe, a cura di Jenn Riley, 2009-2010, http://jennriley.com/metadatamap.

Showers Ben, *Does the library have a role to play in digital humanities?*, "JISC", 2012, http://infteam.jiscinvolve.org/wp/2012/02/23/does-the-library-have-a-role-to-play-in-the-digital-humanities.

Sikos Leslie F., Philp Dean, *Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs*, "Data Science and Engineering", 5 (2020), p. 293-316.

Singhal Amit, *Introducing the Knowledge Graph: Things, Not Strings*, Google Official Blog, May 16, 2012, https://blog.google/products/search/introducing-knowledge-graph-things-not.

Sula Chris Alen, *Digital Humanities and Libraries: A Conceptual Model,* "Journal of Library Administration", 53 (2013), 1, p. 10-26, https://doi.org/10.1080/01930826.2013.756680.

Sweeney Shelley, *Provenance of Archival Materials*, in *Encyclopedia of Library and Information Sciences*, Third Edition, edited by Marcia J. Bates, Mary Niles Maack, vol. VI, p. 4315-4323, Boca Raton Florida, CRC Press, 2010.

Taylor Arlene G., *The organization of information*, Englewood, Libraries Unlimited, 1999.

Testoni Laura, *Digital curation e content curation: due risposte alla complessità dell'infosfera digitale che ci circonda, due sfide per i bibliotecari*, "Bibliotime", 16 (2013), 1.

Thaller Manfred, *Clio – ein datenbankorientiertes system für die historischen Wissenschaften : Fortschreibungsbericht*, "Historical Social Research", 12 (1987), 1, p. 88-91.

Id., *Defining humanities computing methodology*, in *Computing in Humanities Education: A European Perspective*, ACO*HUM, edited by Koenraad De Smedt et al., University of Bergen, HIT-centre, 1999, http://korpus.uib.no/humfak/AcoHum/book.

Tomasi Francesca, *L'Ecdotica e le tecnologie. Tra applicazioni, principi e un esperimento*, "Anuario Lope de Vega. Texto, literatura, cultura", 20 (2014), p. 84-98, http://dx.doi.org/10.5565/rev/anuariolopedevega.84.

Ead., *Edizioni o archivi digitali? Knowledge sites e apporti disciplinari*, in *Edizioni Critiche Digitali. Digital Critical Editions. Edizioni a*

confronto/comparing editions, a cura di Paola Italia, Claudia Bonsi, Roma, Sapienza Università Editrice, 2016, p. 129-136.
Ead., *DIGITAL HUMANITIES AND DIGITAL KNOWLEDGE (DHDK). International second cycle/Master degree*, "Umanistica Digitale", 2 (2018).
Ead., *Digital humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM*, "Aibstudi", 60 (2020), 2, p. 1-15, https://doi.org/10.2426/aibstudi-12068.
Ead., *L'edizione delle carte degli scrittori: l'archivio letterario incontra il (nativo) digitale*, "Biblioteche oggi Trends", 7 (2021), 1, p. 81-91, http://dx.doi.org/10.3302/2421-3810-202101-081-1.
Tomasi Francesca, Ciotti Fabio, Daquino Marilena, Lana Maurizio, *Esplorare semanticamente collezioni culturali: uno studio di fattibilità*, "AIDAinformazioni", 3-4 (2015), p. 125-143, https://doi.org/ 10.4399/97888548899278.
Unsworth John, *What Is Humanities Computing and What Is Not?*, "Jahrbuch für Computerphilologie", 4 (2002), http://hdl.handle.net/2142/157.
Id., *Forms of Attention: Digital Humanities Beyond Representation*, in *The Face of Text: Computer-Assisted Text Analysis in the Humanities*, third conference of the Canadian Symposium on Text Analysis (CaSTA), (McMaster University, November 19-21, 2004), https://johnunsworth.name/FOA.
Valacchi Federico, *I siti web come strumenti per la ricerca archivistica*, "Archivio Storico Italiano", 160 (2002), 3, p. 589-610.
Id., *Things in the World. The integration process of archival descriptions in intercultural systems*, "JLIS.it", 7 (2016), 2, p. 331-67, https://dx.doi.org/10.4403/jlis.it-11529.
Vespasiano da Bisticci, Lettere, edizione a cura di Francesca Tomasi, Bologna, AlmaDL - Università di Bologna, 2013, https://doi.org/10.6092/unibo/vespasianodabisticciletters.
Vespasiano da Bisticci, Lettere, Knowledge Site, versione 3.0, Bologna, Digital Humanities Advanced Research Centre (/DH.arc), Università di Bologna, 2020, http://projects.dharc.unibo.it/vespasiano.
Vivarelli Maurizio, *Digital humanities e culture documentarie: un modello di analisi, valutazione, interpretazione*, "AIB studi", 60 (2020), 3, p. 553-589, https://doi.org/10.2426/aibstudi-12471.
Vogeler, Georg, *Digital Edition of Archival Material - Machine Access to the Content: On the Role of Semantic Web Technologies in Digital*

Scholarly Editions, in *Digitizing Medieval Sources? L'édition en ligne de documents d'archives médiévaux*, edited by Christelle Balouzat-Loubet, Turnhout, Brepols, 2020, p. 37-56, https://doi.org/10.1484/M.ARTEM-EB.5.117327.

Vrandečić Denny, Krötzsch Markus, *Wikidata: A Free Collaborative Knowledgebase*, "Communications of the ACM", 57 (2014), 10, p. 78–85, https://doi.org/10.1145/2629489.

W3C incubator group Library Linked Data, *Datasets, Value vocabularies and Metadata Element Sets*, 2011, https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/.

W3C Working Group, *Best Practices for Publishing Linked Data*, 2014, https://www.w3.org/TR/ld-bp.

Wilkinson Mark D., et al., *The FAIR Guiding Principles for scientific data management and stewardship*, "Scientific Data", 3 (2016), https://www.nature.com/articles/sdata201618.

Zeng Marcia Lei, *Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article*, "El profesional de la información", 28 (2019), 1, https://doi.org/10.3145/epi.2019.ene.03.

Zhang Ying, Liu Shu, Mathews Emilee, *Convergence of digital humanities and digital libraries*, "Library management", 36 (2015), 4-5, p. 362-367, https://doi.org/10.1108/LM-09-2014-0116.

Zorich Diane, Waibel Günter, Erway Ricky, *Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums*, 2008, https://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf.